Identifying Ideological Perspectives in Text and Video

Wei-Hao Lin

CMU-LTI-08-008

October, 2008

Language Technologies Institute
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Alexander G. Hauptmann, Chair
William W. Cohen
Eric P. Xing
Janyce Wiebe, University of Pittsburgh

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Keywords: Ideological Discourse, Statistical Machine Learning, Broadcast Television News

To my parents, my heroes

Abstract

Polarizing opinions about political and social controversies take place commonly in mass and more recently user-generated media. A functional democratic society builds on civic discussions among people holding different beliefs on an issue. However, so far, few computer technologies have been devoted to facilitate mutual understanding, and arguably could have worsened the situation.

We envision a computer system that can automatically understand different ideological viewpoints on an issue and identify biased news stories, blog posts, and television news. Such a computer system will raise news readers' awareness of individual sources' biases and encourage them to seek news stories from different viewpoints.

- Computer understanding of ideological perspectives, however, has been long considered almost impossible. In this thesis, we show that ideology, although very abstract, exhibits a concrete pattern when it is communicated among a group of people who share similar beliefs in written text, spoken text, television news production, and web video folksonomies. This *emphatic* pattern in ideological discourse opens up a new field of automatic ideological analysis, and enables a large amount of ideological text and video to be automatically analyzed.

- We develop a new statistical model, called Joint Topic and Perspective Models, based on the emphatic pattern in ideological discourse. The model combines two essential aspects of ideological discourse: topic matters and ideological biases. The simultaneous inference on topics and ideological emphasis, however, poses a computational challenge. We thus develop an approximate inference algorithm for the model based on variational methods.

- The emphatic pattern in ideological discourse and the Joint Topic and Perspective Model enable many interesting applications in text analysis and multimedia content understanding. At the corpus level, we show that ideological discourse can be reliably distinguished from non-ideological discourse. At the document level, we show that the perspective from which a document is written or a video is produced can be identified with high accuracy. At the sentence level, we extend the model to summarize an ideological document by selecting sentences that strongly express a particular perspective.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Polarizing discussions about political and social issues commonly occur in mass media as well as user-generated media, in television news, newspapers, and blogs. Computer and information technologies so far have not addressed the problem of detecting individual authors' bias and may have actually worsen the situation. Political scientists have argued that news filtering and recommendation technologies prevent readers and viewers from engaging controversial issues and pose a threat to a democratic society (Sunstein, 2007). News aggregation websites (e.g., Google) allow users to pick and choose their favorite news topics and ignore others. The following description from Google News shows that news personalization has been emphasized as its main feature:

> No one can read all the news thats published every day, so why not set up your page to show you the stories that best represent your interests?

Recommendation services such as e.g., Yahoo suggest news articles based solely on users' reading histories.

Computer programs that can automatically identify the perspective from which a text is written or a video is produced will facilitate mutual understanding among people of different cultures and beliefs. Such computer programs can highlight which part of news reports reflects strongly an ideological perspective and can help news viewers become more aware of the bias of individual news networks. Furthermore, computer programs can point viewers to the news stories of opposing perspectives on the same issue from other news networks. News audiences are thus encouraged to consider controversial issues from broader and multiple viewpoints.

We envision a computer system that can automatically identify ideological perspectives of newspaper articles, blogs, radio news reports, television news, and web videos. Such a computer system will work like a GPS in an ideological landscape. It will tell readers and

Figure 1.1: A example news cluster about the United States presidential candidates from Google News as of July 1, 2008

viewers where they are located in an ideological landscape: what ideological perspectives from which a news article is written or a television video is produced. It will also guide readers and viewers to a new ideological view that is different from what they are reading and watching.

Imagine a new news aggregation service that goes beyond grouping news stories from multiple newspapers. Instead of presenting readers with huge cluster of 1,604 news articles about the 2008 United States presidential candidates as shown in Figure 1.1, this new service tells you among these more than a thousand of news articles how many articles are strongly pro-Democratic and pro-Republican.

This new service allows you to sort news stories by how strongly they convey political beliefs. In addition to "sort by relevance" and "sort by date" that are already available on Google News, the new service offers *sort by ideological perspectives*. Instead of scouring through thousands of news articles, you ask the new service to return a list of news articles strongly conveying pro-Republican views:

- Is Democratic presidential candidate Barack Obama becoming a victim of hubris? I ask in the wake of a couple of recent campaign trail developments and a disconcerting personal encounter with the senator's Chicago headquarters.

  – Democratic Hubris? Tuscaloosanews.com, 2008-06-29

- Barack Obama is the liberal Democratic nominee for president. He is not change. There is nothing un-American or disloyal about being a liberal or a conservative. They are two sides of the same coin.

But Sen. Obama is not about change in our politics. Webster's dictionary defines change as, to make radically different; transform. The senator's record on change is not radically different, it is bereft of actions and long on talk.

– McCain Is the Clear Choice for Change, theday.com, 2008-06-29

- Yesterday, Gen. Wesley Clark Attacked McCain's Military Service:

  Gen. Clark: "But he hasn't held executive responsibility. That large squadron in the Navy that he commanded? It wasn't a wartime squadron. He hasn't been there and ordered the bombs to fall. He hasn't seen what it's like when diplomats come in and say, 'I don't know whether we're going to be able to get this point through or not, do you want to take the risk, what about your reputation, how do we handle this publicly?' He hasn't made those points Bob." (CBS' "Face The Nation," 6/29/08)

  – New Politics? Wesley Clark's Attack On McCain's Military Service Demonstrates Obama's 'New Politics' Are Just Words, Republican National Committee, 2008-06-30

- The Ruinous Bequests of the Sixties Most protest movements begin as an organized expression of a legitimate grievance – some perceived societal injustice, perhaps in response to actual governmental or judicial tyranny. If the timing is right and the issues resonate, successful protest movements can flourish and quickly grow into full-fledged revolutions, and revolutions can often degenerate into bloody civil wars.

  – Elect Obama, Destroy America, Roger W. Gardner, The Conservative Voice, 2008-07-01

When a user reads a news article or a blog post, the system will first show the user how biased this text is. Moreover, the system will highlight the paragraphs and sentences that are strongly one-sided. The system will also add hyperlinks on these highly biased sentences and paragraphs, and point the user to those articles that express contrasting ideological views. By "augmenting" news articles (Elo, 1996) and videos with ideological information, readers can become more aware of the ideological perspectives an author takes, and different ideological stances on an issue that might have been missed or ignored.

In this thesis, we study how ideological perspectives are reflected in text and video. Specifically, we are interested in developing a computer system that can automatically identify the ideological perspective from which a text document was written or a video was made.

Our goal is to develop a computer system that can automatically identify highly biased news stories in newspapers, blogs, television news, and web videos. Such a system may increase an audience's awareness of authors' bias in text or video, and can encourage them to seek news stories from contrasting viewpoints. Considering multiple viewpoints could help people make more informed decisions and strengthen democracy. Psychological studies have shown that persuasion with arguments from two sides are more effective than one-sided argument (Hovland, Janis, & Kelley, 1953).

The automatic analysis of ideological perspectives developed in this thesis, combined with other computer technologies, will enable many novel computer applications. Web search engines can use ideology analysis to retrieve documents expressing a particular ideological viewpoint of interest. News aggregation services can use ideology analysis to better present and organize news stories. Instead of presenting a cluster of thousands of news stories, news stories can be grouped by ideological viewpoints. Web advertisement networks can use ideology analysis to target those readers interested in a particular ideological point of view. Content control and web filtering software can incorporate ideology analysis to filter out websites and blogs that express extreme political, social, or religious views that may not be suitable for children.

Classifiers that can automatically identify a web video's ideological perspective will enable video sharing sites to organize videos of various social and political views according to their ideological perspectives, and allow users to subscribe videos based on their personal views. Automatic perspective classifiers will also enable content control or web filtering software to filter out videos expressing extreme political, social, or religious views that may not be suitable for children.

By an ideological perspective, we mean a set of beliefs commonly shared by a group of people (Van Dijk, 1998). Groups whose members share similar goals or face similar problems usually share a set of beliefs that define membership, value judgment, and action. These collective beliefs form an ideology. For example, the Democratic and Republican parties represent two dominant ideological perspectives in the United States politics. Two presidential candidates, John Kerry and George W. Bush, gave the following answers to a question on abortion during the third presidential debate in 2004:

(1.1) Kerry: What is an article of faith for me is not something that I can legislate on somebody who doesn't share that article of faith. I believe that choice is a woman's choice. It's between a woman, God and her doctor. And that's why I support that.

(1.2) Bush: I believe the ideal world is one in which every child is protected in law and welcomed to life. I understand there's great differences on this issue of abortion,

> but I believe reasonable people can come together and put good law in place that will help reduce the number of abortions.

The above examples show that two candidates expressed two very different ideological perspectives on the abortion issue. One candidate takes a so-called “pro-choice” position that values a woman’s choice while the other takes a “pro-life” position that values life of an unborn child.

The difference in framing news events is clearer when we compare news broadcasters across national boundaries, languages, and media. Ideological perspectives are not reflected only in text. Video has been a popular medium for expressing different opinions and value judgments. For example, Figure 1.2 shows how an American broadcaster (NBC) and an Arabic broadcaster (LBC) portray Yasser Arafat’s death in 2004. The two broadcasters’ footages are very different: NBC shows stock footage of Arafat, while LBC shows footage of interviews with general public and the funeral.

(a) The key frames of a television news story from an American news broadcaster, NBC

(b) The key frames of a television news story from an Arabic news broadcaster, LBC

Figure 1.2: The key frames of the television news footages on Yasser Arafat’s death from two broadcasters.

1.1 Ideology

Ideology seems to enjoy the status of “I know it when I see it” as pornography did in the 1964 United States Supreme Court decision (Van Dijk, 1998). Although many scholars in sociology and literature have attempted to define ideology, an exact definition is still elusive.

In this thesis, we follow Van Dijk (1998)'s definition of ideology. Ideology in Van Dijk (1998)'s theory is not narrowly defined as the beliefs of the dominant class as it would be viewed by traditional Marxist sociologists. Instead, Van Dijk (1998) embraces a much broader view and defines ideology as "a set of general beliefs commonly shared by a group of people."

Ideology has been extensively studied in different fields, but Van Dijk (1998)'s ideology theory is unique in combining three core components that have only been studied separately before. The three main components in Van Dijk (1998)s theory are: *cognition*, *society*, and *discourse*. Ideology consists of ideas in peoples minds which are usually studied by cognitive psychologists. Ideology also involves a group membership and value judgment, which are generally studied by sociologists and social scientists. Ideology is not innate knowledge and therefore needs to be reproduced and transmitted through written or spoken discourse. Van Dijk argues that ideology cannot be fully understood unless we not only study all three of these components but also consider the interactions among the three components.

Van Dijk (1998)s ideology theory carefully characterizes the interaction between *cognition* and *society*. The theory holds that ideology is neither merely personal nor just about specific events but general, abstract beliefs. Ideology is not about cultural beliefs that would be shared across otherwise competing social groups. Shared knowledge can sometimes function as ideology, where truth and evaluation criteria defined by one group may be deemed false and not recognized at all by an ideologically opposite group.

Ideology is a set of general beliefs socially shared by a group of people. Since not every social group is defined by ideology, van Dijks theory carefully characterizes those groups that are usually defined by ideology and the social functions that an ideology provides. Groups that commonly exhibit an ideology include socio-economic classes (low-income class vs. high-income class), professional groups (e.g., journalists), social movements (e.g., feminism), religion (e.g., Christianity), ethnics group (e.g., African-Americans), and political parties (e.g., Democrats vs. Republicans).

Ideology, according to Van Dijk (1998), provides the following important social functions:

- Membership: who are we? How do we define ourselves based on characteristics (e.g., gender, race, ethnicity, socio-economic class, age, religion, language, culture, etc.)? Who is our enemy?

- Action: what do we do as a group?

- Value: what do we value most? Of what do we want to convince other people?

- Belief: what do we believe as a group? For example, religious beliefs by a religious group.

- Relationships with other group: where do we stand on certain issues?

- Resources: what (realistic or symbolic) resources do we have and lack? Ideological groups usually protect their resources, or fight for resources they do not have.

Van Dijk (1998)'s ideology theory includes discourse as an indispensable component of an ideology. Since ideology is not based on innate knowledge, people must subscribe to an ideology via written or spoken communication. How the written and spoken discourse advocates, reproduces, and promotes an ideology becomes very important in understanding an ideology and its competing ideologies. Van Dijk (1998) has identified a variety of discourse structures that can carry important functions of ideology, including:

- Syntax: The subject of a sentence reflects what an author holding an ideological perspective wants to emphasize. Pronouns can particularly reflect an ideological groups membership. How do they refer to us vs. them, and in-group (people sharing similar beliefs) vs. out-group (people competing for resources)? The syntactic markers showing politeness (tu and vous in French, and tu in Spanish) also reflect this type of membership relation.

- Semantics: Ideological discourse is persuasive in nature; how historical or social events are portrayed, positively or negatively, clearly reflects an ideology. An ideological group usually praises events that are congruent to their beliefs while they condemn events that are contrary to their beliefs. The lexical choices are classical examples. The word choice between terrorists or "freedom fighters" clearly indicates an opinion about an action considered as very negative and out-group or very positive and in-group. Van Dijk (1998) had found that "variation of lexical items (that is, lexical style) is a major means of ideological expression in discourse."

- Schematic structures: Like the syntactic structures that exist at the sentence level, there are also schematic structures that exist at the discourse level. For example, the words in the title of a news story can strongly indicate a newspapers ideological view on a news event. The description in a background information paragraph, however, usually conveys less of an ideological perspective. Therefore, to understand how an ideology assigns importance to various aspects of an event, it is more significant to consider a portions of text reflect than to know that a discourse generally conveys a particular perspective?

In addition to linguistic and discourse structures, Van Dijk also points out that the context of discourse plays an equally important role in understanding the discourse. The following contexts are explicitly mentioned:

- Domain: Ideological discourse is strongly tied to a domain. The right-to-life (anti-abortion) ideology will only manifest itself in discussions on birth control, sex education, life, and death, but not in discussions of something like organic food.

- Date and time: When a discourse takes place can sometimes add additional meaning and interpretation. A political speech on racism made on Martin Luther Kings Day is not the same as a similar speech on another day.

- Location: Where a discourse is made also adds an important context. A political speech made in a war zone carries a different semantic meaning than a speech given in legislative chambers.

- Social roles: Is the discourse from a middle-class citizen or a company CEO? Is the discourse from a white male or an African-American female?

- Affiliation: Who wrote the discourse? If it is news, which news organization does the reporter belong to?

1.2 Thesis Outline

1.2.1 Modeling Ideological Perspectives

Lexical variations have been identified as a "major means of ideological expression" (Van Dijk, 1998). In expressing a particular ideological perspective, word choices can highly reveal an author's ideological perspective on an issue. "One man's terrorist is another man's freedom fighter." Labeling a group as "terrorists" strongly reveals an author's value judgment and ideological stance (Carruthers, 2000).

We illustrate lexical variations in an ideological text about the Israeli-Palestinian conflict (see Section 3.1.1). There are two groups of authors holding contrasting ideological perspectives (i.e., Israeli vs. Palestinian). We count the words used by each group of authors and show the top 50 most frequent words in Figure 1.3.

Both sides share many words that are highly related to the corpus's topic (i.e., the Israeli-Palestinian conflict): Palestinian, Israeli, political, peace, etc. However, each ideological perspective seems to emphasize (i.e., choosing more frequently and having a bigger word size in Figure 1.3) different sets of words. The Israeli authors seem to more

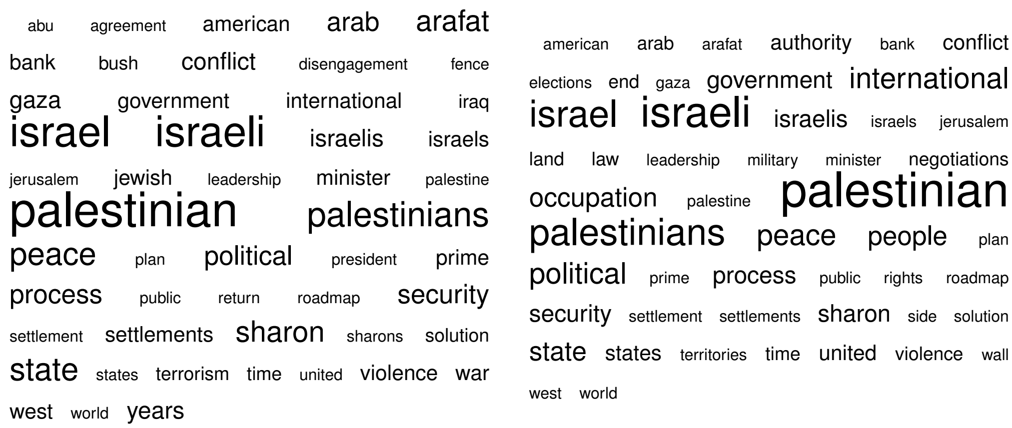


Figure 1.3: The top 50 most frequent words used by the Israeli authors (left) and the Palestinian authors (right) in a document collection about the Israeli-Palestinian conflict. A word's size represents its frequency: the larger, the more frequent.

frequently use disengagement, settlement, and terrorism. In contrast, the Palestinian authors seem to more frequently choose occupation, international, and land. Some words seem to be chosen because they are related to the topic while other words are chosen because of an author's ideological stance.

We thus hypothesize that lexical variations in ideological discourse are attributed to both the ideological text's topic and the author or speaker's ideological point of view. We develop a statistical model to capture the lexical variations in Chapter 4. Word frequency in ideological discourse should be determined by how much a word is related to the text's topic (i.e., *topical*) and how much authors holding a particular ideological perspective emphasize or de-emphasize the word (i.e., *ideological*). A model for ideological discourse should take both topical and ideological aspects into account.

1.2.2 Identifying Ideological Corpus

Ideological perspectives do not always manifest themselves when any two documents are contrasted. Take the following sentences from Reuters news wire as an example:

(1.3) Gold output in the northeast China province of Heilongjiang rose 22.7 pct in 1986 from 1985's level, the New China News Agency said.

(1.4) Exco Chairman Richard Lacy told Reuters the acquisition was being made from

Bank of New York Co Inc, which currently holds a 50.1 pct, and from RMJ partners who hold the remainder.

The above pair of sentences does not exhibit strongly opposing ideological perspectives as do those in the Kerry-Bush answers cited earlier. Rather, as the Reuters indexers did, people would label Example 1.3 as "GOLD" and Example 1.4 as "Acquisition" as two *topics*, not two *perspectives*.

We study the problem of identifying ideological corpora in this thesis. The solution to the problem can address why people perceive different ideological perspectives in the pair of Example 1.1 and Example 1.2, but perceive little ideological perspectives in the pair of Example 1.3 and Example 1.4. In Section 5, based on empirical observation that ideological perspectives are reflected in lexical variations, we take a model-based approach to differentiate ideological corpus from non-ideological text corpus (e.g., a corpus of news articles that cover different news topics but do not express strongly different ideological perspectives.)

The problem of identifying ideological corpus is not only scientifically intriguing, but it also enables us to develop important natural language processing applications that can be used to detect the emergence of contrasting perspectives. Media and political analysts regularly monitor broadcast news, magazines, newspapers, and blogs to see if there are splits in public opinion. The huge number of documents, however, make the task extremely daunting. Therefore, an automated test of different perspectives will be very valuable to information analysts.

The positive experimental results on differentiating ideological discourse from non-ideological discourse in Chapter 5 motivates the development of the statistical model for ideological discourse in Chapter 4. We formalize unique patterns of ideological discourse that are empirically observed in Section 5.1.4 and propose a new statistical model called a Joint Topic and Perspective Model. We show that the proposed model closely captures ideological discourse in various experiments in Section 4.5.

1.2.3 Identifying Ideological Documents

In addition to discovering document collections that contain opposing perspectives, we are interested in identifying documents that are written from a particular perspective. For example, in the context of the Palestinian-Israeli conflict:

(1.5) The inadvertent killing by Israeli forces of Palestinian civilians – usually in the course of shooting at Palestinian terrorists – is considered no different at the moral and ethical level than the deliberate targeting of Israeli civilians by Palestinian suicide bombers.

(1.6) In the first weeks of the Intifada, for example, Palestinian public protests and civilian demonstrations were answered brutally by Israel, which killed tens of unarmed protesters.

Example 1.5 is written from the Israeli perspective; Example 1.6 is written from the Palestinian perspective. Political analysts who follow the development of the Israeli-Palestinian conflict want not only to know that Example 1.5 and Example 1.6 are written from opposing perspectives, but also to look for more documents that are written from a particular perspective of interest.

People knowledgeable about the Israeli-Palestinian conflict can easily identify a document's perspective, but human reviewing is costly when there are huge numbers of documents. Computer programs that can automatically identify a document's perspective will be a valuable tool for people analyzing text from different perspectives. We study the problem of automatically identifying the ideological perspective of a text document in Section 6. Based on the statistical model we develop in Chapter 4, we evaluate the effectiveness of such computer programs in the task of predicting the ideological perspective of a document. More effective computer programs will achieve higher classification accuracy.

1.2.4 Identifying Ideological Sentences

When an issue is discussed from different perspectives, not every sentence strongly reflects the overall perspective of an author. For example, the following sentences are written by one Palestinian and one Israeli, respectively:

(1.7) The Rhodes agreements of 1949 set them as the ceasefire lines between Israel and the Arab states.

(1.8) The green line was drawn up at the Rhodes Armistice talks in 1948-49.

Example 1.7 and Example 1.8 introduce the background of the ceasefire line drawn in 1949, and no explicit perspectives are expressed.

Analysts who sift through large collections of documents are interested in not only quickly retrieving documents of a particular perspective of interest, but also identifying which part of a strongly reflects a perspective. We study the problem of identifying sentences that strongly express an ideological perspective in Section 7.

1.2.5 Identifying Ideological Perspectives in Video

Text is not the only medium in which perspectives are regularly expressed. Video has been a popular medium to convey subjective beliefs and values.

Not every pair of two news clips sharing similar characteristics will exhibit different perspectives. We study the problem of identifying ideological perspectives expressed in video in this thesis.

We focus on broadcast news video in this thesis. Television news has been the predominant way of understanding the world around us. Individual news broadcasters, however, can frame and even mislead audience's understanding about political and social issues. Efron's pioneering study showed that news production involved many decision-making processes, and that news broadcasters' choices varied differently on different political and social issues (Efron, 1972). A dull parade event can be easily manipulated by cameras and suddenly become an event with many participants. Hence, the quote goes, "Cameras don't lie, but liars use cameras." (Berger, 1998)

The bias of individual news broadcasters could heavily shape an audience's views on many social and political issues. A recent study showed that the respondents' main news sources were highly correlated with their misconceptions about the Iraq War (Kull, 2003). 80% of respondents whose primary news source was FOX had one or more misconceptions, while 50% of people whose primary news source was CNN did.

We consider a broadcaster's bias in portraying a news event to be "ideological" because television news production involves a large number of people who share similar social and professional beliefs. We take the definition of ideology as "a set of general beliefs socially shared by a group of people" (Van Dijk, 1998). News production involves many decisions, e.g., what to cover, whom to interview, and what to show on screen. A news broadcaster could consistently introduce bias when reporting political and social issues partly because producers, editors, and reporters collectively make similar decisions based on shared value judgments and beliefs.

Computer and information technologies so far have done little to address the media bias problem, and arguably could have worsened the situation. Many websites (e.g., Google News, My Yahoo, etc.) allow users to pick and choose their favorite news topics. Scholars and political scientists have worried that these news filtering and recommendation technologies prevent readers from engaging controversial issues and pose a threat to a democratic society (Sunstein, 2007).

Video sharing websites such as YouTube, Metacafe, and Imeem have been extremely popular among Internet users. More than three quarters of Internet users in the United States have watched video online. In a single month in 2008, 78.5 million Internet users watched 3.25 billion videos on YouTube. On average, YouTube viewers spend more than one hundred minutes a month watching videos on YouTube (comScore, 2008).

Video sharing websites have also become an important platform for expressing and communicating different views on various social and political issues. In 2008, CNN and YouTube held United States presidential debates in which presidential candidates

Figure 1.4: The key frames of a web video expressing a “pro-life” view on the abortion issue, which is tagged with `prayer`, `pro-life`, and `God`.

Figure 1.5: The key frames of a web video expressing a “pro-choice” view on the abortion issue, which is tagged with `pro`, `choice`, `feminism`, `abortion`, `women`, `rights`, `truth`, `Bush`.

answered questions that were asked and uploaded by YouTube users. In March 2008 YouTube launched YouChoose’08 [1] in which each presidential candidate has his own channel. The accumulative viewership for one presidential candidate as of June 2008 has exceeded 50 million (techPresident, 2008). In addition to politics, many users have authored and uploaded videos expressing their views on social issues.

For example, Figure 1.4 is an example of a “pro-life” web video on the abortion issue[2], while Figure 1.4 is an example of “pro-choice” web video[3].

1.3 Contributions

In this thesis, we make the following contributions to computer understanding of ideological perspectives in text and video.

[1] `http://youtube.com/youchoose`
[2] `http://www.youtube.com/watch?v=TddCILTWNr8`
[3] `http://www.youtube.com/watch?v=oWeXOjsv58c`

- To the best of our knowledge, we are the first to develop *automatic* solutions to the problem of understanding ideological perspectives. In contrast to previous approaches that have heavily relied on manually constructed knowledge base, our statistics-based approaches require almost no human intervention. We provide a low-cost and highly efficient solution to analyzing large number of text and video documents that express different viewpoints on political and social issues.

- We discover an emphatic pattern in ideological discourse. Although ideology is an idea that is difficult to be well defined, the unique emphatic pattern is objectively defined. The emphatic pattern is prevailing. The pattern exhibits itself in text, broadcast news video, and web videos: in words, visual concepts, and user-generated tags.

  The emphatic pattern in ideological discourse accounts for two factors that contribute to word frequency in text or visual concept frequency in video: a topical factor commonly shared by different viewpoints and an ideological factor emphasized or de-emphasized by individual viewpoints. The two factors are mathematically defined, and the contributions of individual factors can be quantified. The pattern points out tangible differences between two viewpoints rather than saying two viewpoints are different but without explaining where the differences lie.

- We propose a statistical model for ideological discourse, called Joint Topic and Perspective Model (jTP). The model is based on the emphatic pattern, and simultaneously captures the topical and ideological factors in an unified statistical model. However, the unification of two numerical factors poses a great challenge on inference. We develop an efficient approximate inference algorithm for jTP using variational methods. Given a training corpus, the model can simultaneously uncover the topical and ideological factors. After the training, the model can predict the ideological viewpoint of a new document using the learned topical and ideological weights.

  jTP provides a human understandable explanation on the difference between two viewpoints (categories). By examining the learned topical and ideological weights, jTP users can clearly understand the underlying assumptions and power of the model. This is very different from many classifiers that focus solely on improving accuracy and provide little explanation on why a model works on a set of data.

- We collect a number of text and video data to evaluate our methods, and these data will be valuable to the Natural Language Processing and Multimedia communities. So far there have been very few publicly available resources for studying ideological discourse. Our annotated resources will enable researchers in Natural Language

Processing and Multimedia who are interested in ideological discourse to quickly setup experiments and evaluate their approaches.

- We attack the problem of understanding ideological perspectives not in single modality but across multiple modalities (text and video). The cross-modal contrast and comparison not only strengthen the findings in the thesis, but also open a new field for studying ideological perspectives. This echoes the cross-disciplinary approach of studying ideology proposed by Van Dijk (1998). Van Dijk (1998) offers many insights by considering ideology across many disciplines: psychology, sociology, and discourse analysis. Similarly, we combine techniques in several computing fields such as Natural Language Processing, Computational Linguistics, Information Retrieval, Multimedia, Image Processing, and Web Sciences and study the ideological discourse in written and spoken documents and in broadcast news videos and user-generated videos.

- Non-computing fields can benefit from the findings and techniques in this thesis. For example, content analysis has been a popular methodology in media studies and sociology (Krippendorff, 1980), but applying content analysis usually requires huge human efforts to code and annotate data. Thus, the analyses are usually done in a small scale. Automatic analysis tools such as our Joint Topic and Perspective Model will enable a much larger, web scale analysis of ideological discourse that has not been done before.

- Our ultimate goal in this thesis is to facilitate mutual understanding among people holding different beliefs on political and social issues. The current Internet technologies allow users to customize their daily news and filter out "irrelevant" news. The filtering technologies inevitably creates many echo chambers: people listen to those who they agree with and people talk to those they like. Our work in this thesis attempts to open a window to individual echo chambers. The automatic ideology analysis can alert Internet users about individual news sources' biases. Whenever they read a biased new article, a blog post, or a YouTube video, our system can alert them. Furthermore, our work builds a bridge between people in different echo chambers. By automatically analyzing a large corpus of documents expressing different views on a political or social issue, our work can point users to documents of a viewpoint different from theirs, and help expose them to a world that they may not have been aware of.

Chapter 2

Literature Review

Automatically identifying ideological perspectives, either in text and video, has not been explored in the literature. However, the research work in this thesis has been heavily influenced by various research problems in the fields of natural language processing, machine learning, and multimedia.

- Probably, the previous work most relevant to this thesis is the early attempt at using computers to model political beliefs in the 1960s. As we will review in Section 2.1, these early attempts relied heavily on *manually* constructed knowledge source, and painted a very pessimistic picture of *automatically* learning ideological beliefs from data. Against this backdrop of pessimistic view on automatically acquiring knowledge that we focus on developing machine learning algorithms that automatically learn ideological perspectives from data.

- An ideology is a set of beliefs in a writer or speaker's mind. Language to convey these inner thoughts is expected to be subjective. Recently, researchers in the field of natural language processing and information retrieval have been interested in the research problem of distinguishing subjective language from objective language. We will summarize the findings from subjectivity analysis with identifying ideological perspectives in Section 2.2.

- Defining membership is one of important social functions that an ideology provides, and language to distinguish us from them usually resorts to "praise us" and "criticize them" (Van Dijk, 1998). The problem of distinguishing positive language from negative language has recently attracted many researchers in the fields of natural language processing and text mining. We will summarize the findings from sentiment analysis in Section 2.3, and contrast sentiment analysis with identifying ideological perspectives.

- Automatically identifying a document's ideological perspective is one of the research goals in this thesis. The problem can seem to be just another text categorization task. However, as we will review in Section 2.4, much previous work in text categorization has focused on identifying the a document's subject matter, and little work has been done to study ideological text.

- Modeling ideological discourse is one of the research goals in this thesis. We borrow techniques and ideas from topic modeling. However, as we will review in Section 2.5, most work on topic modeling has focused on text collection containing multiple (latent) topics (e.g., newspapers), which is very different from ideological text studied in this thesis.

- In addition to text, video has been a popular medium on which ideology is expressed. However, automatically identifying ideological perspectives in video has not been widely studied in the field of multimedia. Some previous work, assuming that a video's perspective is known and need not be identified, has demonstrated the potential impact that an automatic video perspective identifier can make. We will summarize these studies in Section 2.6.

2.1 Computer Modeling of Ideological Beliefs

Computer modeling of belief systems has attracted Artificial Intelligence researchers' attention since the field's inception. Abelson and Carroll (1965) pioneered simulating the belief systems of individuals in computers. The simulation system, known as the Goldwater machine, represented the beliefs of a right-wing politician on foreign policy during the Cold War as a set of English sentences composed of a subject followed by a verb and an object, for example, "Cuba subverts Latin America." Abelson (1973) later extended the simple sentence-based representation to a hierarchical representation. The extended representation, closely following the Schank and Abelson (1977)'s framework of knowledge presentation, distinguished between actions and purposes of actors, captured a sequence of actions for a purpose, and modeled interactions between multiple actors. Carbonell (1978) proposed POLITICS, a simulation system that can interpret a political event described in text from two conflicting ideologies, e.g., conservative and liberal (Carbonell, 1979). POLITICS focused on understanding the goals of actors, and a new structure, goal tree was developed to perform "counter-planning", that is, to thwart other actors from achieving their goals.

The goal of automatically identifying ideological perspectives has not been fully addressed in previous work. Computer simulation in previous work was not an end, but a means of making assumptions about human belief systems explicit. Therefore, early computer simulation programs could neither determine if two text documents expressed conflicting views nor predict the author's ideological perspectives.

Beliefs in previous work were *manually* collected and translated into computer-readable forms, which is very different from our goal of *automatically* learning perspectives from a collection of documents. Previous work takes a top-down approach to modeling beliefs while our approach in this thesis is bottom-up. Manually-constructed knowledge base has been known to suffer from "acquisition bottleneck" (Buchanan et al., 1983) and is difficult to transfer to new domains.

Learning one's attitude toward an issue directly from written or spoken documents has been considered to be impossible. Abelson and Carroll (1965) expressed a very pessimistic view on the possibility of learning beliefs from text without any prior knowledge:

> The simulation of the belief systems of other individuals [other than Goldwater] with very different views is also being contemplated, but this step cannot be undertaken lightly since the paraphrasing procedure [a method of manually representing beliefs in computers] is extremely difficult. One might suppose that fully automatic content analysis methods could be applied to the writings and speeches of public figures, but there is an annoying technical problem which renders this possibility a vain hope.

Instead of subscribing to this view, we believe that statistical modeling allows perspectives to be learned from training documents without human supervision. Part of this thesis's contribution is to show to what degree statistical learning can learn perspectives automatically.

Sack (1994) studied the problem of automatically identifying ideological perspectives based on what *role* an *actor* is portrayed in foreign news reports. An article written from a guerrilla's perspective would be more likely to portray a government (actor) as a victim (role). An article written from a government's perspective would portray guerrillas (actor) as victims (role). They developed a computer system, SpinDoctor, that automatically extracted the actors and their roles in foreign news stories, and determined the article's ideological perspective based on the role-actor analysis. Take the following excerpt from a news story about the Salvadoran Civil War as an example (Sack, 1994, p. 37):

(2.1) On 10 January at 1030, on the 10th anniversary of Radio Venceremos, an FMLN unit commemorated this occasion by ambushing a military vehicle transporting national guardsmen from Sesori to San Miguel. ... A few minutes after the fighting began our troops gained control of the situation.

SpinDoctor analyzed Example 2.1 and output a list of (actor, role) tuples such as (Venceremos, source), (national guardsmen, victim), and (our troops, military). Because the guerrilla radio station, Veneceremos, plays a "source" role, the national guardsmen in the Salvadoran government play "victim" roles, and our troops (referring to FMLN, a guerrilla organization) plays a military role, SpinDoctor classified this article to have the guerrilla's ideological perspective.

Sack (1994)'s work on automatic ideology analysis is different from Abelson and Carroll (1965) and his colleagues' in two important ways. First, Sack (1994) focuses on the surface structure of text, i.e., what is actually written in ideological text. Specifically, Sack (1994) looks at the actors and their roles portrayed in news stories. In contrast, (Abelson & Carroll, 1965; Carbonell, 1978) focus on the deep structure and start from knowledge-intensive modeling of ideological beliefs. Our approach of automatic ideology analysis is closer to Sack (1994)'s bottom-up approach than to Abelson and Carroll (1965) and Carbonell (1978)'s top-down approach. As we argued previously, the top-down approach requires human experts to compile a knowledge base, which is difficult to maintain and adapt to a new issue. In contrast, the bottom-up approach can quickly adapt to a new issue by collecting ideological texts on the new issue. We would not go as far as Anderson (2008)'s claims that in the petabyte age large amount of data can replace science and no scientific theory is needed anymore. However, the benefits from automatically analyzing large amount of data – easy adaptability to a new domain and low human intervention – should not be easily dismissed.

Second, Sack (1994) systematically and objectively evaluates the performance of identifying ideological perspectives. Sack (1994) developed SpinDoctor based on 25 documents from MUC-3 (Message Understanding Conference) (Sundheim, 1991) and evaluated SpinDoctor's performance of identifying a news article's ideological perspective on 75 previously unseen documents. Despite the small data set, Sack (1994) showed that the performance of automatically identifying ideological perspectives could be objectively quantified. In this thesis, we follow a similar methodology to evaluate our method of automatically identifying ideological perspectives, but on a large scale and on a variety of different documents in text and video.

SpinDoctor, however, is seriously limited by the manually-specified patterns of identifying actors and roles in a narrow domain (the Salvadoran Civil War). SpinDoctor requires domain experts to specify possible text patterns for actors (e.g., "national guardsmen" are actors in the government) and for roles (e.g., "X was killed" where X plays a victim role). SpinDoctor matches these patterns against text to identify actors and roles. Similar to the "knowledge bottleneck" in a knowledge-based system, the "pattern bottleneck" prevents these manually specified patterns from generalizing to unseen data. The problem due to the manually specified patterns can be clearly seen in poor performance of SpinDoctor on unseen documents. The manually specified patterns appear to over-fit the training set. The true correct rate is 76% on the training set but only 35% on the testing set. In this thesis, we avoid this problem by automatically acquiring patterns from analyzing large amount of data instead of manually specifying them.

Fortuna, Galleguillos, and Cristianini (2008) explored the problem of identifying media bias. They collected news articles on the Internet and found that the news sources (CNN vs. Al Jazeera) of the news articles on the Middle East can be successfully identified based on word choices using Support Vector Machines (SVM) (Cristianini & Shawe-Taylor, 2000). They identified the words that can best discriminate two news sources using Canonical Correlation Analysis (CCA) (Hotelling, 1936).

In addition to the clearly distinctive methods between Fortuna et al. (2008) and this thesis, there are crucial differences between Fortuna et al. (2008)'s work and this thesis. First, instead of applying two different statistical methods as Fortuna et al. (2008) did, the Joint Topic and Perspective Model (Chapter 4) is a single unified model that can learn to predict an article's ideological slant and uncover discriminating word choices *simultaneously*. Second, the Joint Topic and Perspective Model makes explicit the assumption of the underlying generative process on ideological text. The generative process is modular and can be easily enriched with new linguistic constraints. In contrast, discriminative classifiers such as SVM used in (Fortuna et al., 2008) rely heavily on the feature engineering to encode complex linguistic constraints. Third, Fortuna et al. (2008) focused on the news articles on the Middle East. In contrast, our scope is beyond news sources' bias in regional politics, and we focus on ideological beliefs in general. We evaluate ideological beliefs held by a variety of groups such as Internet users, politicians, and politic pundits on numerous political and social issues such as abortion, the United States politics, gay rights (see Section 3.

2.2 Subjectivity Analysis

Subjectivity analysis refers to analysis of language used in expressing opinions, evaluation, and emotions. J. Wiebe, Wilson, Bruce, Bell, and Martin (2004) has defined subjective language as "language used to express private states in the context of a text or conversation." For example, Example 2.2 contains two highly subjective expressions (in italic)[1]:

(2.2) Although there is only scant possibility of a military conflict in the Taiwan Strait, the Nouvel Observateur said, Beijing's military buildup in its southern coastal

[1]The example is from the document `20010713/00.42.05-29788` in the MPQA corpus (J. Wiebe, Wilson, & Cardie, 2005).

> provinces still makes the region ”*the most heavily armed place in the world* next only to the Middle East,” and *a dangerous flashpoint* in the 21st century.

J. Wiebe et al. (2005) explained “private states” as “a general term that covers opinions, beliefs, thoughts, feelings, emotions, goals, evaluations, and judgments” when they developed an annotation scheme for subjective language. Subjective language in newspapers (e.g., Wall Street Journal (J. Wiebe et al., 2004) and world press (J. Wiebe et al., 2005)) and newsgroups (J. Wiebe et al., 2004) has been annotated and widely studied.

Much research on subjectivity analysis has focused on distinguishing subjective language from objective language. The granularity of annotation on subjective and objective language ranges from documents (J. Wiebe et al., 2004) to sentences (J. Wiebe et al., 2004; Yu & Hatzivassiloglou, 2003) and to expressions (phrases in a sentence) (J. Wiebe et al., 2004). On the surface, the task of determining if a document or a sentence is subjective is very much like a text categorization task (also see Section 2.4).

J. M. Wiebe (1994) and her colleagues pioneered subjectivity analysis and first explored the problem of identifying subjective elements in a sentence. Tracking characters' points of views in novels in (J. M. Wiebe, 1994), however, is not the same as identifying ideological perspectives in this thesis. The points of views in this thesis are ideological (e.g., pro-life vs. pro-choice on the abortion issue) while the points of views in (J. M. Wiebe, 1994) are psychological (e.g., a paragraph in a novel narrated from the first person perspective). Uspensky (1973) illustrated the distinction between ideological perspective and psychological perspective with many examples from novels.

Many features have been identified to be useful for distinguishing subjective language from objective language. A *bag-of-words* representation (or unigram), commonly used in the fields of text categorization and information retrieval (Lewis, 1998), has been shown to be very effective in identifying subjective documents and sentences (Yu & Hatzivassiloglou, 2003). J. Wiebe et al. (2004) have identified many useful features beyond unigram, including *hapax legomenon* (words occurring only once in a text collection), n-gram of lexicon and part-of-speech tuples, and words that are collocated with subjective verbs and adjectives. Riloff, Wiebe, and Wilson (2003) showed that subjective nouns could improve subjectivity classification. Riloff and Wiebe (2003) extracted subjective patterns (e.g., ⟨Subject⟩ complained, where ⟨Subject⟩ is the subject of a sentence) from a large collection of unannotated documents.

Subjectivity classifiers have many applications. Researchers in information retrieval have started to study the problem of retrieving blog posts that are not only relevant but also opinioned (Ounis, Rijke, Macdonald, Mishne, & Soboroff, 2006).

There are crucial differences between subjectivity analysis and identifying ideological perspectives. First, the research goals are different. Ideological text can be very subjective because writers and speakers want to convey their thoughts, and can be considered as a special kind of subjective text. Subjectivity analysis aims to learn features that discriminate subjective language from objective language. In contrast, identifying ideological perspectives aims to learn features that distinguish one ideological perspective from a contrasting perspective.

Second, ideology can be expressed in both subjective and objective language. Therefore, labeling a sentence as subjective is not enough to determine if the sentence conveys, if any, ideological perspectives. Example 2.3 shows that seemingly objective expressions can convey ideological information. Although the report stated an attorney general's resignation objectively, readers can deduce the Dann's ideological perspective based on his political party.

(2.3) Dann, a Democrat, resigned as attorney general amid impeachment action during an investigation into sexual harassment and mismanagement of the office[2].

Although the research goals of subjectivity analysis and identifying ideological perspectives are different, subjectivity analysis can be incorporated to improve the performance of identifying ideological perspectives. Because ideological text is mostly subjective by its nature, one hypothesis is that ideological perspectives are largely expressed in subjective language and less in objective language. If the hypothesis holds true, by excluding or down-weighting the objective part of a document we may improve the performance of identifying a document's ideological perspective.

2.3 Sentiment Analysis

The problem of distinguishing positive sentiment from negative sentiment has attracted interests in the fields of natural language processing and text mining. After subjective documents are identified or when starting from presumably subjective documents (e.g., movie (Pang, Lee, & Vaithyanathan, 2002) or product reviews (Morinaga, Yamanishi, Tateishi, & Fukushima, 2002; Dave, Lawrence, & Pennock, 2003)), sentiment classifiers automatically categorize words (Turney & Littman, 2003; Beineke, Hastie, & Vaithyanathan, 2004), sentences (Yu & Hatzivassiloglou, 2003), or documents (Dave et al., 2003; Pang et al., 2002) into expressing positive sentiment (praise) or negative sentiment (criticism). Pang and Lee (2008) gave a comprehensive review of the challenges and techniques in sentiment analysis.

Researchers have identified many features useful in identifying positive and negative sentiments. Unigrams (or, in general, n-grams), previously shown to be very effective

[2] Josh Sweight, *Local man to help GOP pick attorney general candidate*, May 28, 2008, Middletown Journal.

in text categorization, perform well in sentiment classification (Pang et al., 2002). In addition to n-grams, researchers have evaluated many features on sentiment classification such as part-of-speech and lexical tuples (Dave et al., 2003), sub-strings (Dave et al., 2003), and collocations (Dave et al., 2003). Subjective words have been used as seed words to discover collocations (Turney & Littman, 2003; Yu & Hatzivassiloglou, 2003; Dave et al., 2003). A unique feature for sentiment classification is negation (i.e., replacing "not good" as a single token "NOTgood") (Dave et al., 2003; Pang et al., 2002), which greatly changes the polarity of a word and can not be captured in unigrams.

Pang and Lee (2004) showed that subjectivity analysis (also see Section 2.2) can further improve sentiment classification. Combining multiple knowledge sources has also been shown to improve sentiment classification. Mullen and Collier (2004) fused Turney and Littman (2003)'s polarity value, WordNet, and Nasukawa and Yi (2003)'s idea of focusing on sentiment targeting a specific subject (e.g., the singer in a CD review). V. Ng, Dasgupta, and Arifin (2006) combined unigrams with n-grams, manually choosing subjectives and objective words.

The research goals of sentiment analysis and identifying ideological perspectives are very different. Sentiment analysis is about identifying language used to express positive and negative opinions, and this is not the same as identifying one ideological perspective from a contrasting ideological perspective. Ideological perspectives are reflected in many ways other than sentiments.

The difference between sentiment analysis and identifying ideological perspectives can be shown in the following examples. A word "criticism", ostensibly expressing negative sentiment, can be used to convey contrasting ideological perspectives. In the context of the Israeli-Palestinian conflict, the "criticism" in Example 2.4 is toward Arafat and thus conveys the Israeli point of view. In contrast, the "criticism" in Example 2.5 is from 144 countries including the US and thus conveys the contrasting, Palestinian point of view.

(2.4) In this regard, too, harsh *criticism* of Yasir Arafat was to be expected, given PM Sharon's broad success in discrediting the Palestinian leader as a terrorist and a pathological liar. .

(2.5) Israel, on the other hand, has continued building the apartheid/separation wall despite the United Nations vote of 144 countries condemning the wall, and in spite of the public American *criticism*. .

Therefore, identifying its sentiment is not enough to distinguish a sentence's ideological perspective.

Sentiment analysis can be incorporated to improve the performance of identifying ideological perspective in text. Ideological text is expected to be subjective and interspersed with positive sentiments toward people sharing similar beliefs and negative sentiments toward others (Van Dijk, 1998). We may thus hypothesize that ideological perspectives are partially reflected in the opinion holder and target of positive and negative sentiments. If the hypothesis holds true, the performance of identifying ideological perspectives can be improved by identifying the opinion holder and target of positive and negative sentiments.

2.4 Text Categorization

Text documents can be classified into pre-defined categories with high accuracy. The problem of text categorization has been extensively studied, from comprehensive comparison between competing classifiers (Yang & Liu, 1999; Sebastiani, 2002) to feature selection (Yang & Pedersen, 1997), to new classification algorithms (Joachims, 1998; McCallum & Nigam, 1998), and to utilization of unlabeled data (Nigam, McCallum, Thrun, & Mitchell, 2000).

We borrow many techniques and evaluation methodology from text categorization. The most popular and successful choice for text representation is a bag-of-words representation. Each document is represented as a vector whose coordinates are the count of a term within the document, i.e., term frequency (TF), and the inverted count of a term appearing in multiple documents, i.e., inverted document frequency (IDF). The bag-of-words representation ignores word order and does not utilize rich information in syntax and semantics. It makes strong assumptions that words are independent from each other, which is not true in natural languages. However, the bag-of-words representation has been shown very effective in many natural language processing tasks, including text categorization (Sebastiani, 2002) and information retrieval (Lewis, 1998).

Identifying a document's ideological perspective can be regarded as a special kind of text classification task. An ideological perspective classifier determines if a document belongs to one ideological perspective or the contrasting ideological perspective. However, so far the popular categories in text categorization have been subject matters (e.g., news topics in Reuters newswire [3]). It is not very clear how successfully the text categorization approach will perform for "ideological" text. Although research on text categorization has started to move from news topic classification to emails (Klimt & Yang, 2004) and hierarchical classification (Lewis, Yang, Rose, & Li, 2004), very few studies focus on ideological documents. Notable text classification tasks not on news topics include genre detection (Kessler, Nunberg, & Schütze, 1997), subjectivity detection (see Section 2.2), sentiment detection (see Section 2.3), and authorship attribution (Mosteller & Wallace, 1984).

[3] http://www.daviddlewis.com/resources/testcollections/reuters21578/

2.5 Topic Modeling

Research on topics models (Blei, Ng, & Jordan, 2003; Griffiths & Steyvers, 2004; Rosen-Zvi, Griffths, Steyvers, & Smyth, 2004; McCallum, Corrada-Emmanuel, & Wang, 2004) (also see a survey paper (Steyvers & Griffiths, In Press)) shows promising results on recovering the latent structure in topical documents. They provide a solid statistical learning foundation for us to further investigate the interaction between topics and ideological perspectives. Unfortunately, similar to text classification (Section 2.4), most work in topic modeling has focused on news articles (e.g., TREC AP[4]) and academic publications, and little work focuses on ideological text.

There have been studies that model beyond topics (e.g., modeling authors (Rosen-Zvi et al., 2004)). However, we are interested in modeling lexical variation collectively from multiple authors sharing similar beliefs, not lexical variations due to individual authors. When we collect ideological discourse data (see Chapter 3), we ensure that documents are contributed by multiple authors, not individuals.

2.6 Ideology in Video

So far there has been very little work in the field of multimedia on automatically identifying a news video's ideological perspective. To the best of our knowledge, our work is the first to automatically identify a news video's ideological perspective.

The most relevant work includes a video symmetrization system based on a viewer's attitude on war (Bocconi & Nack, 2004) and multimedia art installations that promote mutual understanding between people holding different ideological viewpoints. VOX POPULI (Bocconi & Nack, 2004) is a computer system that can make a documentary from a pool of interview clips based on the viewer's position on an issue, e.g., "Iraq War." Minions (Ireson, 2004) is an interactive art installation that confronts visitors with videos from two religious perspectives, Christianity and Islam. Arango (2004) displays a multimedia art, Vanishing Point, that shows us how mainstream news media in industrialized countries give uneven coverage of countries around the world. "Terminal Time" (Mateas, Vanouse, & Domike, 2000; Mateas, 2002) is a video generation system that automatically generates ideologically-biased documentaries based on Carbonell (1978)'s ideology goal trees. However, they all assume that videos' ideological perspectives are known.

Besides the art installation and video generation work, very few works in the field of multimedia have studied the problem of identifying different perspectives in video. In previous work, the ideological perspective of a video in previous work is either assumed to

[4] http://www.daviddlewis.com/resources/testcollections/trecap/

be known or manually labeled. Manual annotation makes it almost impossible to analyze large number of videos. Instead we are interested in developing *automatic* methods to identify the perspectives of videos.

There has been research on linking stories on the same topic across news sources (or "topic detection" (Allan, 2002)), using cues in key frame images (Zhai & Shah, 2005), visual concept (Zhang, Lin, Chang, & Smith, 2004), or near-duplicates (X. Wu, Hauptmann, & Ngo, 2007) to cluster news on the same event across different news channels. We linked news stories on the same topic based on automatic speech recognition transcriptions. This keyword-based topic detection approach was simple but not perfect. The perspective identification performance shown later in Section 6.2.3.2 could have been further improved if we had used better topic detection techniques.

Linking stories across multiple news channels is a necessary component in a television news system. To contrast how individual news sources select and compose footages, we need an efficient and effective way of selecting all news videos on the same news event from multiple news channels. Visual similarity between two news stories is shown to be of moderate help, and text similarity (from closed captions or ASR transcripts) contributes much more to the success of linking stories in broadcast news (Zhai & Shah, 2005).

Chapter 3

Experimental Data

We prepared text corpora consisting of documents that were written or spoken from contrasting perspectives. The first corpus, bitterlemons, contains written documents that were written from an Israeli or a Palestinian perspective (Section 3.1.1)[1]. The second corpus, 2004 Presidential Debates, consists of spoken documents that were spoken by Kerry or Bush in the 2004 Presidential Debates (Section 3.1.2).

To test how well our methods can distinguish document collections of contrasting perspectives from documents of no perspectives, we need a corpus of documents that are commonly regarded as different from each other in any way but in "perspectives." We focus on a particular difference, *topicality*, and choose a corpus, Reuters-21578, that contains news stories in different topics (Section 3.1.3).

3.1 Text Data

3.1.1 Bitterlemons

The bitterlemons corpus consists of the articles published on the website `http://bitterlemons.org/`. The website is set up to "contribute to mutual understanding [between Palestinians and Israelis] through the open exchange of ideas."[2] Every week, an issue about the Israeli-Palestinian conflict is selected for discussion (e.g., "Disengagement: unilateral or coordinated?"), and a Palestinian editor and an Israeli editor contribute one article each addressing the issue. In addition, the Israeli and Palestinian editors invite one Israeli and one Palestinian to express their views on the issue (sometimes in the form of an interview),

[1] The author of this thesis would like to thank Theresa Wilson for first mentioning this website to the author.

[2] `http://www.bitterlemons.org/about/about.html`

resulting in a total of four articles in a weekly edition. We collected a total of 594 articles published on the website from late 2001 to early 2005. The distribution of documents and sentences are listed in Table 3.1.

	Palestinian	Israeli
Written by editors	148	149
Written by guests	149	148
Total number of documents	297	297
Average document length	740.4	816.1
Number of sentences	8963	9640

Table 3.1: The basic statistics of the corpus

We chose the bitterlemons.org website for two reasons. First, each article is already labeled as either Palestinian or Israeli by the editors. Second, the bitterlemons corpus enables us to test the generalizability of the proposed methods in a very realistic setting: training on articles written by a small number of writers (two editors) and testing on articles from a much larger group of writers (more than 200 different guests).

We removed metadata from all articles, including edition numbers, publication dates, topics, titles, author names and biographies. We used OpenNLP Tools[3] to automatically decide sentence boundaries and reduced word variants using the Porter stemming algorithm (Porter, 1980).

To test if the ratio of subjective sentences to objective sentences would help distinguish one perspective from the other, we estimated the subjectivity of each sentence using an automatic subjective sentence classifier (Riloff & Wiebe, 2003). We found that 65.6% of Palestinian sentences and 66.2% of Israeli sentences were classified as subjective. The almost equivalent percentages of subjective sentences in the two perspectives support that a perspective is largely expressed in subjective language, but that the amount of subjective sentences in a document is not necessarily indicative of its perspective. One perspective is not necessarily more subjective than the other perspective.

3.1.2 2004 Presidential Debates

The 2004 Presidential Debates corpus consists of the spoken transcripts of three Bush-Kerry debates in 2004. The transcripts are from the Commission on Presidential Debates[4]. We segmented the transcripts according to the speaker tags in the transcripts. Each spoken

[3]http://sourceforge.net/projects/opennlp/
[4]http://www.debates.org/pages/debtrans.html

document was either an answer to a question or a rebuttal. The words from moderators were discarded.

3.1.3 Reuters-21578

The Reuters-21578 corpus[5] is newswire from Reuters in 1987. Reuters-21578 is one of the most common testbeds for (topical) text categorization. Each document is classified into none, one, or more of the 135 categories (e.g., "Mergers/Acquisitions" and "U.S. Dollars"). The number of documents in each category is not evenly distributed (median 9.0, mean 105.9). To perform reliable statistical estimation, we consider only the seven most frequent categories (more than 500 documents) in our experiments: ACQ, CRUDE, EARN, GRAIN, INTEREST, MONEY-FX, and TRADE (in the Reuters codes).

The number of documents, average document length, and vocabulary size of three text corpora are summarized in Table 3.2.

3.2 Video Data

3.2.1 TRECVID 2005 Video Archive

We evaluated the method to identify differing ideological perspectives on a broadcast news video archive from the 2005 TREC Video Evaluation (TRECVID) (Over, Ianeva, Kraaij, & Smeaton, 2005). Since 2001, TRECVID has been a public forum for researchers to evaluate their video processing systems on a common, large video collection. The evaluation tasks include shot detection, high-level feature extraction, video retrieval, and video summarization. The TRECVID'05 video collection is comprised of broadcast news programs recorded in late 2004 in three languages: Arabic, Chinese, and English. Every video is a one-hour or half-hour news program. The number of videos in each language is in Table 3.3.

We used the official shot boundaries that the TRECVID organizer, NIST, provided for the TRECVID 2005 participants. We ran an in-house story segmentation program to detect news story boundaries (A. G. Hauptmann et al., 2005), resulting in 4436 news stories. The story segmentation program detected a news story's boundary using cues such as an anchor's presence, commercials, color coherence, and average story length. We removed anchor and commercial shots because they contained mostly talking heads and conveyed little ideological perspective (in visual content).

[5] http://www.ics.uci.edu/~kdd/databases/reuters21578/reuters21578.html

Corpus	Subset	$\lvert\mathcal{D}\rvert$	$\lvert\bar{d}\rvert$	V
Bitterlemons	Palestinian	290	748.7	10309
	Israeli	303	822.4	11668
	Palestinian Editor	144	636.2	6294
	Palestinian Guest	146	859.6	8661
	Israel Editor	152	819.4	8512
	Israel Guest	151	825.5	8812
2004 Presidential Debates	Kerry	178	124.7	2554
	Bush	176	107.8	2393
	1st Kerry	33	216.3	1274
	1st Bush	41	155.3	1195
	2nd Kerry	73	103.8	1472
	2nd Bush	75	89.0	1333
	3rd Kerry	72	104.0	1408
	3rd Bush	60	98.8	1281
Reuters-21578	ACQ	2448	124.7	14293
	CRUDE	634	214.7	9009
	EARN	3987	81.0	12430
	GRAIN	628	183.0	8236
	INTEREST	513	176.3	6056
	MONEY-FX	801	197.9	8162
	TRADE	551	255.3	8175

Table 3.2: The number of documents $|\mathcal{D}|$, average document length $|\bar{d}|$, and vocabulary size V of three text corpora.

We collected ten news events in late 2004 and news videos covering these news events. We made sure the news events in Table 3.4 had been covered by broadcasters in more than one language. A news story covered a news event if a news event's keywords were mentioned in the video's English automatic speech recognition (ASR) transcripts. NIST provided English translations for non-English news programs. ASR transcripts were only used for linking stories on the same news event. LSCOM annotators did not use ASR transcripts and made judgments solely based on visual content.

Language	Duration	Channels
Arabic	33	LBC
Chinese	52	CCTV, NTDTV
English	73	CNN, NBC, MSNBC

Table 3.3: The channels and the duration of broadcast news video (in hours) in each language in the TRECVID'05 video archive.

News Event	Stories
Iraq War	231
United States presidential election	114
Arafat's health	308
Ukrainian presidential election	11
AIDS	21
Afghanistan situation	42
Tel Aviv suicide bomb	2
Powell's resignation	45
Iranian nuclear weapon	46
North Korea nuclear issue	51

Table 3.4: The number of television news stories on the ten news events in late 2004

3.2.2 LSCOM

We used visual concepts annotation from the Large-Scale Concept Ontology for Multimedia (LSCOM) v1.0 (Kennedy & Hauptmann, 2006). The LSCOM annotations consisted of the presence of each of the 449 LSCOM visual concepts in every video shot of the TRECVID 2005 videos. There are a total of 689064 annotations for the 61901 shots, and the median number of annotations per shot is 10. Examples of images labeled with LSCOM concepts are shown in Figure 3.1.

We first conducted the experiments using the LSCOM annotations, and later replaced manual annotations with predictions from empirically trained concept classifiers. Using manual annotations is equivalent to using very accurate concept classifiers. Given that the state-of-the-art classifiers for most visual concepts are far from perfect, why would we start from manual annotations and assuming perfect concept classifiers? It is because manual annotations allow us to test the idea of measuring similarity in visual concept without being confounded by the poor accuracy of the concept classifiers.

3.2.3 YouTube Tags

We collected web videos expressing opinions on various political and social issues from YouTube[6]. To identify web videos expressing a particular ideological perspective on an issue, we selected "code words" for each ideological perspective and submitted the code words as queries to YouTube. All of the returned web videos were labeled as expressing the particular ideological perspective. For example, the query words for the "pro-life" perspective on the abortion issue are "pro-life" and "abortion."

	Issue	View 1	View 2
1	Abortion	pro-life	pro-choice
2	Democratic party primary election in 2008	pro-Hillary	pro-Obama
3	Gay rights	pro-gay	anti-gay
4	Global warming	supporter	skeptic
5	Illegal immigrants to the United States	Legalization	Deportation
6	Iraq War	pro-war	anti-war
7	Israeli-Palestinian conflict	pro-Israeli	pro-Palestinian
8	United States politics	pro-Democratic	pro-Republican

Table 3.5: Eight political and social issues and their two main ideological perspectives

We downloaded web videos and associated tags for 16 ideological views in May 2008 (two main ideological perspectives for eight issues), as listed in Table 3.5. Tags are keywords voluntarily added by authors or uploaders[7]. The total number of downloaded videos and associated tags are shown in Table 3.6. Note that the number of downloaded videos is

[6] http://www.youtube.com/.

[7] http://www.google.com/support/youtube/bin/answer.py?hl=en&answer=55769

	total videos	total tags	vocabulary
1	2850	30525	4982
2	1063	13215	2315
3	1729	18301	4620
4	2408	27999	4949
5	2445	25820	4693
6	2145	25766	4634
7	1975	22794	4435
8	2849	34222	6999

Table 3.6: The total number of downloaded web videos, the total number of tags, and the vocabulary size (the number of unique tags) for each issue

less than the total number of videos returned by YouTube due to the limit on the maximum number of search results in YouTube APIs.

We assume that web videos containing the "code words" of an ideological perspective in tags or descriptions convey the particular view, but this assumption may not be true. YouTube and many web video search engines have not been designed to retrieve videos expressing opinions on an issue, let alone to retrieve videos expressing a particular ideological view using keywords. Moreover, a web video may contain the code words of an ideological perspective in titles, descriptions, or tags without expressing any opinions on an issue. For example, a news clip tagged with "pro-choice" may simply report a group of pro-choice activists in a protest and not strongly express a so-called pro-choice point of view on the abortion issue. In other words, these YouTube tags are noisy labels for machine learning tasks, and the results using these tags should be considered as lower bound. The performance can be further improved if the tags are manually cleaned.

(a) Vehicle, Armed_Person, Sky, Outdoor, Desert, Armored_Vehicles, Daytime_Outdoor, Machine_-Guns, Tanks, Weapons, Ground_Vehicles

(b) Walking, Hill, Male_Person, Civilian_Person, Standing, Standing, Logos_Full_Screen, Adult, Adult, Walking_Running, Sky, Animal, Person, Outdoor, Clouds, Daytime_Outdoor, Flags, Group, Powerplants, Suits

Figure 3.1: The key frames of two shots from TRECVID'05 and their LSCOM annotations.

Chapter 4

Joint Topic and Perspective Models

We propose a statistical model for ideological discourse. The model associates *topical* and *ideological* weights to each word in the vocabulary. Topical weights represent how frequently a word is chosen because of a document's topic regardless of an author or speaker's ideological perspective. Ideological weights, dependent on an author or speaker's ideological perspective on an issue, modulate topical weights to increase or decrease a word's frequency.

We illustrate the interaction between topical and ideological weights in a three-word simplex in Figure 4.1. A point T represents topical weights about a specific topic. Suppose that authors holding a particular perspective emphasize the word w_3, while authors holding the contrasting perspective emphasize the word w_1. Ideological weights associated with the first perspective will move a multinomial distribution's parameter from T to a new position V_1, which is more likely to generate w_3 than T is. Similarly, ideological weights associated with the second perspective will move the multinomial distribution's parameter from T to V_2, which is more likely to generate w_1 than T is.

We present the Joint Topic and Perspective Model in the domain of ideological discourse in written or spoken text, but later we will show that the model is applicable beyond text. We use terms such as "word", "vocabulary", and "document" to describe the model in this chapter, but these terms should be interpreted broadly. A "document" can be a television news video or a web video. For television news videos, "visual concept" (see Section 6.2) is equivalent to "word", and for web videos, a tag added by users (see Section 6.3) is equivalent to "word."

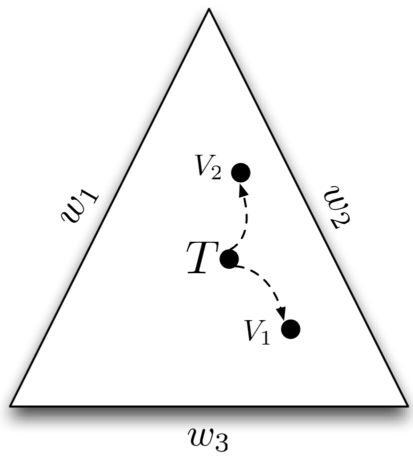


Figure 4.1: A three-word simplex illustrates the main idea behind the Joint Topic and Perspective Model. T denotes the proportion of the three words (i.e., topical weights) that are chosen for a particular topic. V_1 denotes the proportion of the three words after the topical weights are modulated by authors or speakers holding one particular ideological perspective; V_2 denotes the proportion of the weights modulated by authors or speakers holding the other particular set of ideological beliefs.

4.1 Model Specification

Formally, the Joint Topic and Perspective Model assumes the following generative process for ideological discourse:

$$\begin{aligned} P_d &\sim \text{Bernoulli}(\pi), d = 1, \ldots, D \\ W_{d,n} | P_d = v &\sim \text{Multinomial}(\beta_v), n = 1, \ldots, N_d \\ \beta_v^w &= \frac{\exp(\tau^w \times \phi_v^w)}{\sum_{w'} \exp(\tau^{w'} \times \phi_v^{w'})}, v = 1, \ldots, V \\ \tau &\sim \text{N}(\mu_\tau, \Sigma_\tau) \\ \phi_v &\sim \text{N}(\mu_\phi, \Sigma_\phi). \end{aligned} \tag{4.1}$$

The graphical representation of the Joint Topic and Perspective Model is shown in Figure 4.2. The ideological perspective P_d from which the d-th document in a collection is written is sampled from a Bernoulli distribution with a parameter π. In this thesis, we

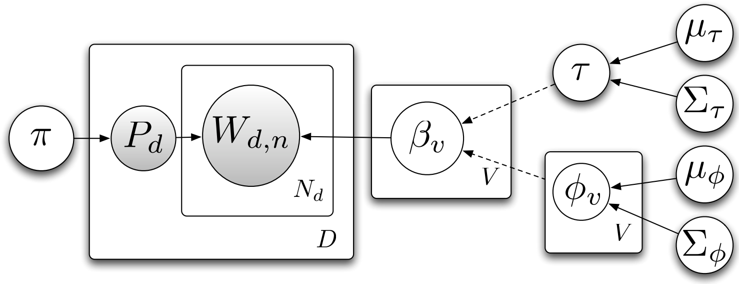


Figure 4.2: The Joint Topic and Perspective Model in a graphical model representation (see Section 4.1 for details). A dashed line denotes a deterministic relation between parent and children nodes.

focus on bipolar ideological perspectives, that is, those political and social issues with two major perspectives ($V = 2$). There are a total of D documents in the collection. The frequency of the n-th word in the d-th document, $W_{d,n}$, is dependent on the ideological perspective of the document's author P_d, and is sampled from a multinomial distribution of a parameter β. There are a total of N_d words in the d-th document.

The multinomial parameter, β_v^w, subscripted by an ideological perspective v and superscripted by the w-th word in the vocabulary, consists of two parts: a topical weight τ^w and ideological weights $\{\phi_v^w\}$. Every word is associated with one topical weight τ^w and two ideological weights ϕ_1^w and ϕ_2^w. β is an auxiliary variable and is deterministically determined by (unobserved) topical and ideological weights.

τ represents the *topical* weights and is sampled from a multivariate normal distribution of a mean vector μ_τ and a variance matrix Σ_τ. ϕ_v represents the *ideological* weights and is assumed to be sampled from a multivariate normal distribution of a mean vector μ_ϕ and a variance matrix Σ_τ. Both τ and ϕ_v are real vectors of the same dimensionality as the total number of words, i.e., vocabulary size.

The prior distributions for topical and ideological weights are normal distributions. Topical weights are modulated by ideological weights through a multiplicative relationship. Therefore, a word with an ideological weight $\phi = 1$ means that the word is not emphasized or de-emphasized. All the weights are normalized through a logistic transformation in (4.1).

The parameters of the model, denoted as Θ, include: π, μ_τ, Σ_τ, μ_ϕ, and Σ_ϕ.

We call this model a Joint Topic and Perspective Model (jTP). To distinguish the model in this section from various extensions in later sections, we sometimes refer to this basic model as the Joint Topic and Perspective Model 1 (jTP1).

4.2 Variational Inference

The quantities of most interest in the Joint Topic and Perspective Model are (unobserved) words' topical weights τ and ideological weights $\{\phi_v\}$. Given a set of D documents on a particular topic with differing ideological perspectives $\{P_d\}$, the joint posterior probability distribution of the topical and ideological weights under the Joint Topic and Perspective Model is

$$\begin{aligned}
&P(\tau, \{\phi_v\} | \{W_{d,n}\}, \{P_d\}; \Theta) \\
\propto &P(\tau|\mu_\tau, \Sigma_\tau) \prod_v P(\phi_v|\mu_\phi, \Sigma_\phi) \prod_{d=1}^{D} P(P_d|\pi) \prod_{n=1}^{N_d} P(W_{d,n}|P_d, \tau, \{\phi_v\}) \\
= &\mathrm{N}(\tau|\mu_\tau, \Sigma_\tau) \prod_v \mathrm{N}(\phi_v|\mu_\phi, \Sigma_\phi) \prod_d \mathrm{Bernoulli}(P_d|\pi) \prod_n \mathrm{Multinomial}(W_{d,n}|P_d, \beta),
\end{aligned}$$

where $N(\cdot)$, $\mathrm{Bernoulli}(\cdot)$, and $\mathrm{Multinomial}(\cdot)$ are the probability density functions of multivariate normal, Bernoulli, and multinomial distributions, respectively.

The joint posterior probability distribution of τ and $\{\phi_v\}$, however, is not computationally tractable because of the non-conjugacy between normal and multinomial distributions. We thus approximate the posterior probability distribution using a variational method (Jordan, Ghahramani, Jaakkola, & Saul, 1999) and estimate the parameters using variational expectation maximization (Attias, 2000). Based on the Generalized Mean Field Theorem (GMF) (Xing, Jordan, & Russell, 2003), we approximate the joint posterior probability distribution of τ and $\{\phi_v\}$ as the product of individual functions of τ and ϕ_v:

$$P(\tau, \{\phi_v\} | \{P_d\}, \{W_{d,n}\}; \Theta) \approx q_\tau(\tau) \prod_v q_{\phi_v}(\phi_v), \tag{4.2}$$

where $q_\tau(\tau)$ and $q_{\phi_v}(\phi_v)$ are the posterior probabilities of the topical and ideological weights conditioned on observed data and GMF messages received from nodes on its Markov blanket.

Specifically, q_ϕ is defined as follows,

$$\begin{aligned}
q_\tau(\tau) = &P(\tau | \{W_{d,n}\}, \{P_d\}, \{\langle\phi_v\rangle\}; \Theta) &(4.3)\\
\propto &P(\tau|\mu_\tau, \Sigma_\tau) \prod_v P(\langle\phi_v\rangle|\mu_\phi, \Sigma_\phi) P(\{W_{d,n}\}|\tau, \{\langle\phi_v\rangle\}, \{P_d\}) &(4.4)\\
\propto &\,\mathrm{N}(\tau|\mu_\tau, \Sigma_\tau)\, \mathrm{Multinomial}(\{W_{d,n}\}|\{P_d\}, \tau, \{\langle\phi_v\rangle\}), &(4.5)
\end{aligned}$$

where $\langle \phi_v \rangle$ denotes the GMF message based on $q_{\phi_v}(\cdot)$. From (4.4) to (4.5) we dropped the terms unrelated to τ.

Calculating the GMF message for τ from (4.5) is computationally intractable because of the non-conjugacy between multivariate normal and multinomial distributions. We followed the approach in (Xing, 2005), and made a Laplace approximation of (4.5). We first represented the word likelihood $\{W_{d,n}\}$ as the following exponential form:

$$P(\{W_{d,n}\}|\{P_d\}, \tau, \{\langle \phi_v \rangle\}) = \exp\left(\sum_v n_v(\langle \phi_v \rangle \bullet \tau) - \sum_v n_v^T \mathbf{1} C(\langle \phi_v \rangle \bullet \tau)\right) \tag{4.6}$$

where $\bullet$ is element-wise vector product, n_v is a word count vector under the ideological perspective v, $\mathbf{1}$ is a column vector of one's, and C function is defined as follows,

$$C(x) = \log\left(1 + \sum_{p=1}^{P} \exp x_p\right), \tag{4.7}$$

where P is the dimensionality of the vector x.

We expanded C using Taylor series to the second order around $\hat{x}$ as follows,

$$C(x) \approx C(\hat{x}) + \nabla C(x)(x - \hat{x}) + \frac{1}{2}(x - \hat{x})^T H(\hat{x})(x - \hat{x}),$$

where ∇C is the gradient of C, and H is the Hessian matrix of C. We set $\hat{x}$ as $\langle \tau \rangle^{(t-1)} \bullet \langle \phi_v \rangle$. The superscript denotes the GMF message in the $t-1$ (i.e., previous) iteration.

The gradient of $C(x)$ in (4.7) with respect to the i-th component can be derived as follows,

$$\frac{\partial C}{\partial x_i} = \frac{\exp x_i}{1 + \sum_{p=1}^{P} \exp x_p}.$$

The Hessian matrix of $C(x)$ is defined as follows,

$$H(x) = \begin{bmatrix} \frac{\partial C}{\partial x_1 \partial x_1} & \cdots & \frac{\partial C}{\partial x_1 \partial x_P} \\ \vdots & \ddots & \vdots \\ \frac{\partial C}{\partial x_P \partial x_1} & \cdots & \frac{\partial C}{\partial x_P \partial x_P} \end{bmatrix},$$

where individual elements can be derived as follows,

$$\begin{aligned} \frac{\partial C}{\partial x_i \partial x_j} &= \frac{-\exp x_i \exp x_j}{\left(1 + \sum_{p=1}^{P-1} \exp x_p\right)^2} \\ \frac{\partial C}{\partial x_i \partial x_i} &= \frac{\exp x_i \left(1 + \sum_{p=1, p \neq i}^{P-1} \exp x_p\right)}{\left(1 + \sum_{p=1}^{P-1} \exp x_p\right)^2} \end{aligned}$$

Finally, we plugged the second-order Taylor expansion of C back to (4.5) and rearranged terms about τ. We obtained the multivariate normal approximation of $q_\tau(\cdot)$

$$q_\tau(\tau) = P(\tau|\{W_{d,n}\}, \{P_d\}, \{\langle\phi_v\rangle\}; \Theta) \tag{4.8}$$

$$\propto P(\tau|\mu_\tau, \Sigma_\tau) \prod_v P(\langle\phi_v\rangle|\mu_\phi, \Sigma_\phi) P(\{W_{d,n}\}|\tau, \{\langle\phi_v\rangle\}, \{P_d\}) \tag{4.9}$$

$$\approx \mathrm{N}(\tau|\mu^*, \Sigma^*), \tag{4.10}$$

where μ^* and Σ^* can be derived as follows:

$$\Sigma^* = \left(\Sigma_\tau^{-1} + \sum_v n_v^T \mathbf{1} \langle\phi_v\rangle \downarrow H(\hat{\tau} \bullet \langle\phi_v\rangle) \rightarrow \langle\phi_v\rangle\right)^{-1}$$

$$\mu^* = \Sigma^* \left(\Sigma_\tau^{-1}\mu_\tau + \sum_v n_v \bullet \langle\phi_v\rangle - \sum_v n_v^T \mathbf{1} \nabla C(\hat{\tau} \bullet \langle\phi_v\rangle) \bullet \langle\phi_v\rangle \right.$$
$$\left. + \sum_v n_v^T \mathbf{1} \langle\phi_v\rangle \bullet (H(\hat{\tau} \bullet \langle\phi_v\rangle)(\hat{\tau} \bullet \langle\phi_v\rangle))\right),$$

where $\downarrow$ is a column-wise vector-matrix product, and $\rightarrow$ is a row-wise vector-matrix product. The Laplace approximation for the logistic-normal prior has been shown to be tight (Ahmed & Xing, 2007).

q_{ϕ_v} in (4.2) can be approximated in a similar fashion as a multivariate normal distribution with a mean vector $\mu^\dagger$ and a variance matrix $\Sigma^\dagger$ as follows:

$$\Sigma^\dagger = \left(\Sigma_\phi^{-1} + n_v^T \mathbf{1} \langle\tau\rangle \downarrow H(\langle\tau\rangle \bullet \hat{\phi}_v) \rightarrow \langle\tau\rangle\right)^{-1}$$

$$\mu^\dagger = \Sigma^\dagger \left(\Sigma_\phi^{-1}\mu_\phi + n_v \bullet \langle\tau\rangle - n_v^T \mathbf{1} \nabla C(\langle\tau\rangle \bullet \hat{\phi}_v) \bullet \langle\tau\rangle \right.$$
$$\left. + n_v^T \mathbf{1} \langle\tau\rangle \bullet (H(\langle\tau\rangle \bullet \hat{\phi}_v)(\langle\tau\rangle \bullet \hat{\phi}_v))\right),$$

where we set $\hat{\phi}_v$ as $\langle\phi_v\rangle^{(t-1)}$.

In E-step, we have a message passing loop and iterate over the q functions in (4.2) until convergence. We monitor the change in the auxiliary variable β and stop when the absolute change is smaller than a threshold. In M-step, π can be easily maximized by taking the sample mean of $\{P_d\}$. We monitor the data likelihood and stop the variational EM loop when the change of data likelihood is less than a threshold.

4.3 Identifiability

The Joint Topic and Perspective Model as specified in Section, 4.1 is, however, not identifiable. There are multiple assignments of topical weights τ and ideological weights $\{P_d\}$ that would result in exactly the same data likelihood. Therefore, topic and ideological weights estimated from data may be incomparable.

The first source of un-identifiability is due to the multiplicative relationship between τ and ϕ_v. We can easily multiply a constant to τ^w and divide ϕ_v^w by the same constant, and the auxiliary variable β stays the same.

The second source of un-identifiability comes from the sum-to-one constraint in the multinomial distribution's parameter β. Given a vocabulary $\mathcal{W}$, we have only $|\mathcal{W}| - 1$ number of free parameters for τ and $\{P_d\}$. Allowing $|\mathcal{W}|$ number of free parameters makes topical and ideological weights unidentifiable.

To solve the un-identifiability issue, we fix the following parameters τ^1, $\{\phi_1^w\}$, and ϕ_v^1. We fix the corner points of τ and $\{\phi_v\}$ (i.e., τ^1 and $\{\phi_v^1\}$) to be 0 and 1, respectively. We choose the first ideological perspective as a base and fix its ideological weights ϕ_1^w to be one for all words.

4.4 Classifying Ideological Perspective

We can apply the Joint Topic and Perspective Model to predict the ideological perspective from which a document is written. We first fit the Joint Topic and Perspective Model on a training corpus $\{W_{d,n}\}$ each of which is labeled with its ideological perspective $\{P_d\}$. Given a document $\{\tilde{W}_n\}$ of an unknown ideological perspective, we use the model with the learned parameters to predict its ideological perspective $\tilde{P}_d$.

Formally, to predict a document's ideological perspective is to calculate the following conditional probability:

$$
\begin{aligned}
& P(\tilde{P}_d | \{P_d\}, \{W_{d,n}\}, \{\tilde{W}_n\}; \Theta) \\
= & \int \int P(\tilde{P}_d, \tau, \{\phi_v\} | \{P_d\}, \{W_{d,n}\}, \{\tilde{W}_n\}) d\tau d\phi_v \\
= & \int \int P(\{\phi_v\}, \tau | \{P_d\}, \{W_{d,n}\}, \{\tilde{W}_n\}; \Theta) \\
& \quad P(\tilde{P}_d | \{\tilde{W}_n\}, \tau, \{\phi_v\}; \Theta) d\tau d\phi_v
\end{aligned} \tag{4.11}
$$

As the predictive probability distribution in 4.11 is not computationally tractable, we approximate it by plugging in the expected values of τ and $\{P_d\}$ obtained in Section 4.2.

4.5 Experiments

4.5.1 Synthetic Data

We first evaluated the model on synthetic data. We fixed the values of the topical and ideological weights, and generated synthetic data according to the generative process in Section 4.1. We tested if the variational inference algorithm for the Joint Topic and Perspective Model in Section 4.2 successfully converges. More importantly, we tested if the variational inference algorithm can correctly recover the true topical and ideological weights that generated the synthetic data.

Specifically, we generated the synthetic data with a three-word vocabulary and topical weights $\tau = (2, 2, 1)$, shown as $\circ$ in the simplex in Figure 4.3. We then simulated different degrees to which authors holding two contrasting ideological beliefs emphasized words. We let the first perspective emphasize w_2 ($\phi_1 = (1, 1+p, 0)$) and let the second perspective emphasize w_1 ($\phi_2 = (1 + p, 1, 0)$. There was no emphasis on w_3. We varied the value of p ($p = 0.1, 0.3, 0.5$) and plotted the corresponding auxiliary variable β in the simplex in Figure 4.3. We generated the equivalent number of documents for each ideological perspective and varied the number of documents from 10 to 1000.

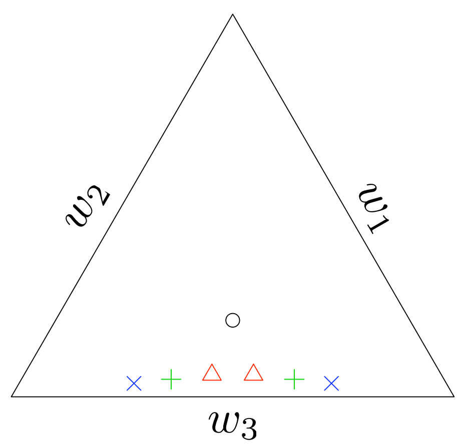


Figure 4.3: We generated synthetic data with a three-word vocabulary. The $\circ$ indicates the value of the true topical weight τ. $\triangle$, $+$, and $\times$ are β after τ is modulated by different ideological weights $\{\phi_v\}$.

We evaluate how closely the variational inference algorithm recovered the true topical and ideological weights by measuring the maximal absolute difference between the true β (based on the true topical weights τ and ideological weights $\{\phi_v\}$) and the estimated $\hat{\beta}$ (using the expected topical weights $\langle\tau\rangle$ and ideological weights $\{\langle\phi_v\rangle\}$ returned by the variational inference algorithm).

The simulation results in Figure 4.4 suggested that the variational inference algorithm for the Joint Topic and Perspective Model is valid and effective. Although the variational inference algorithm was based on Laplace approximation, the inference algorithm recovered the true weights very closely, the absolute difference between true β and estimated $\hat{\beta}$ was small and close to zero.

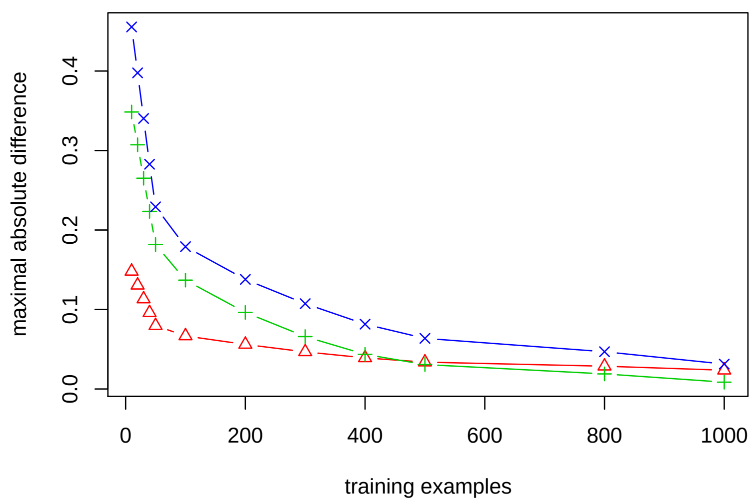


Figure 4.4: The experimental results of recovering true topical and ideological weights. The x axis is the number of training examples, and the y axis is the maximal absolute difference between true β and estimated $\hat{\beta}$. The smaller the difference, the better. The curves in $\triangle$, $+$, and $\times$ correspond to the three different ideological weights in Figure 4.3.

We conducted one more simulation experiment with vocabulary sizes larger than three in Figure 4.4. We varied the vocabulary size from 2^2 to 2^{10}. We first randomly sampled topical weights and ideological weights. For each document, we first sampled a document perspective P_d, and sampled 1000 words. We varied the number of documents from 2^5 to 2^{15}. We measured the relative error of individual β parameters, and calculated the average error over all words in the vocabulary. The relative error is the absolute difference between the true value of β and the learned expected value of β divided by the absolute value of β.

The simulation results with large vocabulary in Figure 4.5 shows that the variational

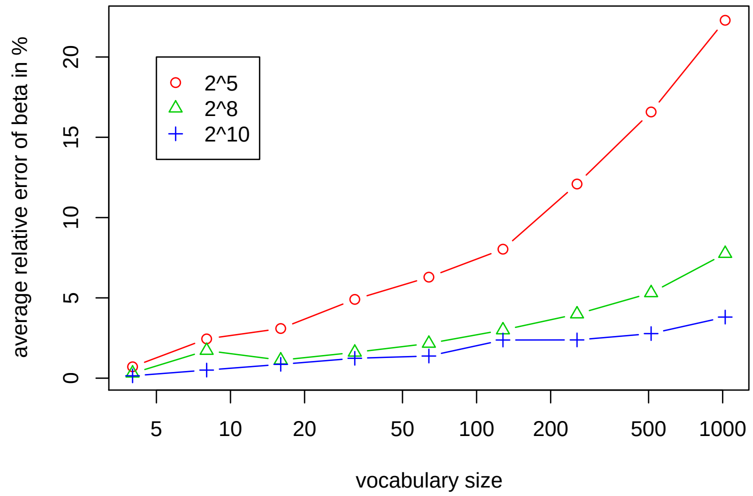


Figure 4.5: The relative error of recovering β parameters (the y axis) of the Joint Topic and Perspective Model under different vocabulary sizes (the x axis). The three curves indicate different numbers of generated documents, from 2^5 to 2^{10}. The y axis is in percentage, and the x axis is in logarithmic scale.

inference algorithms can recover the parameters correctly when vocabulary size is large. A large vocabulary size, however, requires more training data to achieve low error. For example, the fewest number of training documents (2^5, red circle in Figure 4.5) can recover β parameters on average under 5% relative error when vocabulary size is smaller than 2^5 (32). The error rate increases as the vocabulary size increases, which is not completely a surprise.

4.5.2 Ideological Discourse

We evaluated the Joint Topic and Perspective Model on two ideological discourses. The first corpus, bitterlemons, is comprised of editorials written by the Israeli and Palestinian authors on the Israeli-Palestinian conflict. The second corpus, presidential debates, is comprised of spoken words from the Democratic and Republican presidential candidates in 2000 and 2004.

The bitterlemons corpus consists of the articles published on the website `http://bitterlemons.org/`. The website is set up to "contribute to mutual understanding [between Palestinians and Israelis] through the open exchange of ideas."[1] Every week an issue about the Israeli-Palestinian conflict is selected for discussion (e.g., "Disengagement: unilateral or coordinated?"). The website editors have labeled the ideological perspective of each published article. The bitterlemons corpus has been previously used to learn individual perspectives (Lin, Wilson, Wiebe, & Hauptmann, 2006), but it was based on naive Bayes models and did not simultaneously model topics and perspectives.

The 2000 and 2004 presidential debates corpus consists of the spoken transcripts of six presidential debates and two vice-presidential debates in 2000 and 2004. We downloaded the speech transcripts from the American Presidency Project[2]. The speech transcripts came with speaker tags, and we segmented the transcripts into spoken documents according to speakers. Each spoken document was either an answer to a question or a rebuttal. We discarded the words from moderators, audience, and reporters.

We chose these two corpora for the following reasons. First, the two corpora contain political discourse with strong ideological differences. The bitterlemons corpus contains the Israeli and the Palestinian perspectives; the presidential debates corpus the Republican and Democratic perspectives. Second, they are from multiple authors or speakers. There are more than 200 different authors in the bitterlemons corpus; there are two Republican candidates and four Democratic candidates. We are more interested in ideological discourse expressing socially shared beliefs, and less interested in individual authors or candidates' personal beliefs. Third, we selected one written text and one spoken text to test how our model behaves on different communication media.

We removed metadata that might reveal an author or speaker's ideological stance but were not actually written or spoken. We removed the publication dates, titles, author names and biographies in the bitterlemons corpus. We removed speaker tags, debate dates, and location in the presidential debates corpus. Our tokenizer removed contractions, possessives, and cases.

The bitterlemons corpus consists of 594 documents. There are a total of 462,308 words, and the vocabulary size is 14,197. They are 302 documents written by the Israeli authors and 292 documents written by the Palestinian authors. The presidential debates corpus consists of 1232 spoken documents. There are a total of 122,056 words, and the vocabulary size is 16,995. There are 235 spoken documents from the Republican candidates, and 214 spoken documents from the Democratic candidates.

[1] http://www.bitterlemons.org/about/about.html
[2] http://www.presidency.ucsb.edu/debates.php

4.5.3 Topical and Ideological Weights

We fitted the Joint Topic and Perspective Model on two text corpora, and the results are shown in Figure 4.6 and Figure 4.7 in color text clouds[3]. Text clouds represent a word's frequency in size. The larger a word's size, the more frequently the word appears in a text collection. Here we have matched a word's size with its topical weight τ.

Text clouds have been a popular method of summarizing tags and topics on the Internet. Del.icio.us uses tag clouds to help users organize and bookmark. Flickr offers tag clouds as a search interface. Amazon uses tag clouds (called "concordance") to summarize the words in a book. Most tag clouds encode one dimension of information in a word's size. In this thesis we encode two dimensions of information in a word's size and color (Rivadeneira, Gruen, Muller, & Millen, 2007), and succinctly summarize the topical and ideological weights in a color text cloud. Many Eyes allows users to encode two dimensions of information by juxtaposing two words. Compared to Many Eyes, our color text clouds are more spatially efficient. Given the same space, our color text clouds can present more words than those of Many Eyes.

To show a word's ideological weight, we painted words in color shades. We assigned each ideological perspective a color (red or blue). A word's color is determined by which perspective uses a word more frequently than the other. Color shades gradually change from pure colors (strong emphasis) to light gray (no emphasis). The degree of emphasis is measured by how extreme a word's ideological weight ϕ is away from one (i.e., no emphasis). Color text clouds allow us to present three kinds of information at the same time: words, their topical weights, and ideological weights.

Let us focus on the words of large topical weights learned from the bitterlemons corpus (i.e., words in large sizes in Figure 4.6). The word with the largest topical weight is "Palestinian", followed by "Israeli", "Palestinians", "peace", and "political". The topical weights learned by the Joint Topic and Perspective Model clearly match our expectation from the discussions about the Israeli-Palestinian conflict. Words in large sizes summarize well what the bitterlemons corpus is about.

Similarly, a brief glance over the words with large topical weights learned from the presidential debates corpus (i.e., words in large sizes in Figure 4.7) clearly tells us the debates' topic. Words of large topical weights capture what American politics is about (e.g., "people", "president", "America", "government") and specific political and social issues (e.g., "Iraq", "taxes", "Medicare"). Although not every word with large topical weights is attributed to the text's topic, e.g., "im" ("I'm" after contraction is removed) occurred frequently because of the spoken nature of debate speeches, the majority of words with large topical weights appear to convey what the two text collections are about.

[3]We omitted the words of low topical and ideological weights due to space limit.

Figure 4.6: Visualizing the topical and ideological weights learned by the joint topic and perspective model from the bitterlemons corpus (see Section 4.5.3). A word's size is positively correlated its topical weight. Red: words emphasized more by the Israeli authors. Blue: words emphasized more by the Palestinian authors.

Now let us turn our attention to words' ideological weights ϕ, i.e., color shades in Figure 4.6. The word "terrorism", followed by "terrorist", is highly emphasized by the Israeli authors and is painted pure red. "Terrorist" is a word that clearly reveals an author's attitude toward the other group's violent behavior. Many words with large ideological weights can be categorized into the ideological discourse structures that have previously been manually identified by researchers in discourse analysis (Van Dijk, 1998):

- Membership: Who are we and who belongs to us? "Jews" and "Jewish" are used more frequently by the Israeli authors than the Palestinian authors. "Washington" is used more frequently by the Republican candidates than Democratic candidates.

- Activities: What do we do as a group? "Unilateral", "disengagement", and "withdrawal" are used more frequently by the Israeli authors than the Palestinian authors. "Resistance" is used more frequently by the Palestinian authors than the Israeli authors.

- Goals: What is our group's goal? (Stop confiscating) "land", "independent", and (opposing settlement) “expansion” are used more frequently by the Palestinian authors than the Israeli authors.

- Values: How do we see ourselves? What do we think is important? “Occupation” and (human) “rights” are used more frequently by the Palestinian authors than the Israeli authors. “Schools”, “environment”, and “middle” “class” are used more frequently by the Democratic candidates than the Republican candidates. “Freedom” and “free” are used more frequently by the Republican candidates.

- Position and Relations: what are our position and our relation to other groups? “Jordan” and “Arafats” (after removing contraction of “Arafat’s”) are used more frequently by the Israeli authors than by the Palestinian authors.

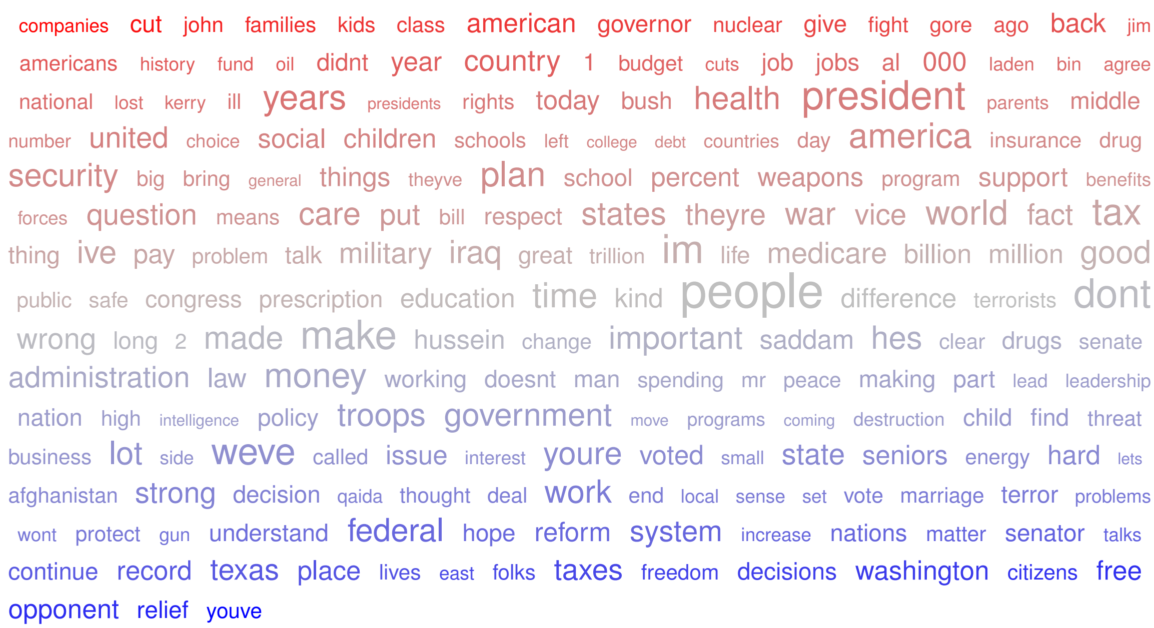


Figure 4.7: Visualizing the topical weights and ideological weights learned by the Joint Topic and Perspective Model from the presidential debates corpus (see Section 4.5.3). A word’s size is positively correlated with its topical weight. Red: words emphasized by the Democratic candidates. Blue: words emphasized by the Republican candidates.

We do not intend to give a detailed analysis of the political discourse in the Israeli-Palestinian conflict and in American politics. We do, however, want to point out that the Joint Topic and Perspective Model seems to “discover” words that play important roles in ideological discourse. The results not only support the hypothesis that ideology is greatly reflected in an author or speaker's lexical choices, but also suggest that the Joint Topic and Perspective Model closely captures the lexical variations.

Political scientists and media analysts can formulate research questions based on the uncovered topical and ideological weights, such as: what are the important topics in a text collection? What words are emphasized or de-emphasized by which group? How strongly are they emphasized? In what context are they emphasized? The Joint Topic and Perspective Model can thus become a valuable tool to explore ideological discourse.

Our results also point out the model's weaknesses. First, a bag-of-words representation is convenient but fails to capture many linguistic phenomena in political discourse. "Relief" is used to represent tax relief, marriage penalty relief, and humanitarian relief. Proper nouns (e.g., "West Bank" in the bitterlemons corpus and "Al Quida" in the presidential debates corpus) are broken into multiple pieces. N-grams do not solve all the problems. The discourse function of the verb "increase" depends much on the context. A presidential candidate can "increase" legitimacy, profit, or defense, and single words cannot distinguish them.

4.5.4 Prediction

We evaluated how well the Joint Topic and Perspective Model predicted words from unseen ideological discourse in terms of perplexity on a held-out set. Perplexity has been a popular metric to assess how well a statistical language model generalizes (Manning & Schütze, 1999). A model generalizes well if it achieves lower perplexity. We choose unigram as a baseline. Unigram is a special case of the Joint Topic and Perspective Model that assumes *no* lexical variations are due to an author or speaker's ideological perspective (i.e., fixing all $\{\phi_v\}$ to one).

Perplexity is defined as the exponential of the negative log word likelihood with respect to a model normalized by the total number of words:

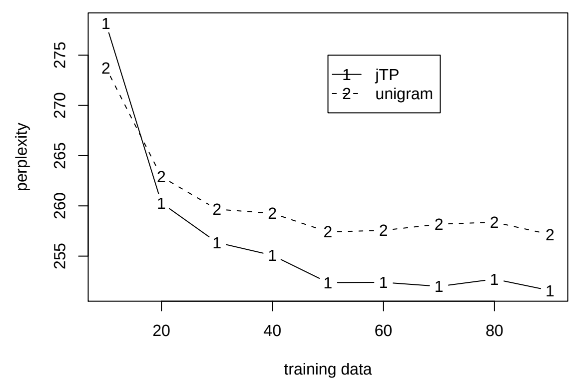
$$\exp\left(\frac{-\log P(\{W_{d,n}\}|\{P_d\};\Theta)}{\sum_d N_d}\right)$$

We could integrate out topical and ideological weights to calculate the predictive probability $P(\{W_{d,n}\}|\{P_d\};\Theta)$:

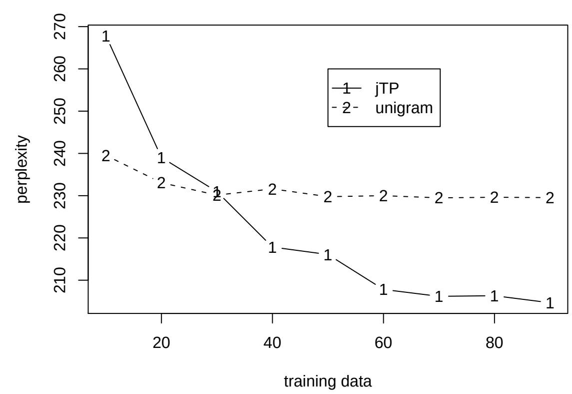
$$P(\{W_{d,n}\}|\{P_d\};\Theta) = \int\int\prod_{d=1}^{D}\prod_{n=1}^{N_d} P(W_{d,n}|P_d)d\tau d\phi_v.$$

Instead, we approximated the predictive probability by plugging in the point estimates of τ and ϕ_v from the variational inference algorithm.

For each corpus, we varied the number of training documents from 10% to 90% of the documents, and measured perplexity on the remaining 10% held-out set. The results are shown in Figure 4.8. We can clearly see that the Joint Topic and Perspective Model reduces perplexity on both corpora. The results strongly support the hypothesis that ideological perspectives are reflected in lexical variations. Only when ideology is reflected in lexical variations can we observe the perplexity reduction from the Joint Topic and Perspective Model. The results also suggest that the Joint Topic and Perspective Model closely captures the lexical variations due to an author or speaker's ideological perspective.



(a) bitterlemons



(b) presidential debates

Figure 4.8: The Joint Topic and Perspective Model reduces perplexity on a held-out set.

Chapter 5

Identifying Ideological Perspectives at the Corpus Level

In this chapter, we first study how to automatically determine if two document collections are written from different ideological perspectives in Section 5.1. By an ideological perspective we mean a point of view, for example, from the perspective of Democrats or Republicans. We propose a test of different perspectives based on distribution divergence between the statistical models of two collections. We apply the same methodology to on both text and video corpora. Experimental results show that the test can successfully distinguish document collections of different perspectives from other types of collections, both in text and video.

Television news has been the predominant way of understanding the world around us, but individual news broadcasters can frame audience's understanding about political and social issues. We aim to develop a computer system that can automatically identify highly biased television news and encourage audience to seek news stories from contrasting viewpoints. However, it is not clear at all if computers can identify news videos produced by broadcasters with differing ideological beliefs. In Section 5.2 we developed a method of identifying differing ideological perspectives based on a large-scale visual concept ontology, and the experimental results were promising.

5.1 Differentiating Ideological Text

We take a model-based approach to develop a computational definition of different perspectives. We first develop statistical models for the two document collections, $\mathcal{A}$ and $\mathcal{B}$, and then measure the degree of contrast by calculating the "distance" between $\mathcal{A}$ and $\mathcal{B}$. Section 5.1.1 describes how document collections are statistically modeled and how distribution difference is estimated. The document corpora are described in Section 3.1. In Section 5.1.2, we evaluate how effective the test of different perspectives based on statistical distribution is. The experimental results show that the distribution divergence can successfully separate document collections of different perspectives from other kinds of collection pairs. We also investigate if the pattern of distribution difference is due to personal writing or speaking styles in Section 5.1.3.

5.1.1 Measuring Difference between Text Corpora

We take a model-based approach to measure to what degree, if any, two document collections are different. A document is represented as a point in a V-dimensional space, where V is vocabulary size. Each coordinate is the frequency of the word within the document, that is, term frequency. Although vector representation, commonly known as a bag of words, is oversimplified and ignores rich syntactic and semantic structures, more sophisticated representation requires more data to obtain reliable models. In practice, bag-of-word representation has been very effective in many tasks, including text categorization (Sebastiani, 2002) and information retrieval (Lewis, 1998).

We assume that a collection of N documents, $y_1, y_2, \ldots, y_N$ is generated by the following sampling process,

$$\begin{aligned} \theta &\sim \text{Dirichlet}(\alpha) \\ y_i &\sim \text{Multinomial}(n_i, \theta). \end{aligned}$$

We first sample a V-dimensional vector θ from a Dirichlet prior distribution with a hyperparameter α, and then sample a document y_i repeated from a multinomial distribution conditioned on the parameter θ, where n_i is the length of the ith document in the collection and assumed to be known and fixed.

We update our knowledge about θ with the information in the documents by Bayes' Theorem,

$$\begin{aligned} p(\theta|\mathcal{A}) &= \frac{p(\mathcal{A}|\theta)p(\theta)}{p(\mathcal{A})} \\ &= \text{Dirichlet}(\theta|\alpha + \sum_{y_i \in \mathcal{A}} y_i). \end{aligned}$$

The posterior distribution $p(\theta|\cdot)$ is again a Dirichlet distribution because a Dirichlet distribution is a conjugate prior for a multinomial distribution.

We are interested in comparing θ of two document collections $\mathcal{A}$ and $\mathcal{B}$, but they are not directly observable. How can we measure the difference between two posterior distributions $p(\theta|\mathcal{A})$ and $p(\theta|\mathcal{B})$? One way to measure the difference between distributions is

Kullback-Leibler (KL) divergence (Kullback & Leibler, 1951; Cover & Thomas, 1991), defined as follows:

$$D(p(\theta|\mathcal{A})||p(\theta|\mathcal{B})) = \int p(\theta|\mathcal{A}) \log \frac{p(\theta|\mathcal{A})}{p(\theta|\mathcal{B})} d\theta. \tag{5.1}$$

KL divergence is asymmetric, and we take the average of $D(P||Q)$ and $D(Q||P)$ as the (symmetric) distance between P and Q, as known as Jensen-Shannon divergence or Information Radius (Manning & Schütze, 1999).

Directly calculating KL divergence according to (5.1) involves high-dimensional integral and is difficult. Alternatively, we approximate the value of KL divergence using Monte Carlo methods as follows:

1. Sample $\theta_1, \theta_2, \ldots, \theta_M$ from $\mathrm{Dirichlet}(\theta|\alpha + \sum_{y_i \in \mathcal{A}} y_i)$.
2. Return $\hat{D} = \frac{1}{M} \sum_{i=1}^{M} \log \frac{p(\theta_i|\mathcal{A})}{p(\theta_i|\mathcal{B})}$ as a Monte Carlo estimate of $D(p(\theta|\mathcal{A})||p(\theta|\mathcal{B}))$.

Algorithms of sampling from a Dirichlet distribution can be found in (Ripley, 1987). As $M \to \infty$, the Monte Carlo estimate will converge to true KL divergence by the Law of Large Numbers.

5.1.2 Experiments on Differentiating Ideological Text

A test of different perspectives is acute only when it can draw distinctions among document collection pairs of different perspectives, document collection pairs of the same perspective, and others. We evaluate the test of different perspectives and apply it to the following four types of document collection pairs $(\mathcal{A}, \mathcal{B})$:

Different Perspectives (DP) $\mathcal{A}$ and $\mathcal{B}$ are written from different perspectives. For example, $\mathcal{A}$ is written from the Palestinian perspective and $\mathcal{B}$ is written from the Israeli perspective.

Same Perspective (SP) $\mathcal{A}$ and $\mathcal{B}$ are written from the same perspective. For example, $\mathcal{A}$ and $\mathcal{B}$ consist of the words spoken by Kerry.

Different Topics (DT) $\mathcal{A}$ and $\mathcal{B}$ are written on different topics. For example, $\mathcal{A}$ is about acquisition and $\mathcal{B}$ is about crude oil.

Same Topic (ST) $\mathcal{A}$ and $\mathcal{B}$ are written on the same topic. For example, $\mathcal{A}$ and $\mathcal{B}$ are both about earnings.

The effectiveness of the test of different perspectives can be measured by how the distribution divergence of DP document collection pairs is separated from the distribution divergence of SP, DT, and ST document collection pairs. The less the overlap, the more acute the test of different perspectives.

Note that the ST + SP condition is the same as SP because our definition of ideology already implicitly includes the requirement of the same topic (ST). Similarly, ST + DP is the same as DP. The DT + SP condition is impossible under our definition of ideology. Ideology must be about the same topic (ST), and it is impossible to express the same perspective (SP) on different topics (DT). Finally, the DT + DP condition is impossible under our definition of ideology. Different ideological perspectives (DP) do not make sense when two corpora are already about different topics (DT)

To account for the considerable variation in the number of words and vocabulary size across corpora (see Table 3.2), we normalize the total number of words in a document collection to be the same K, and consider only the top $C\%$ frequent words in the document collection pair in the estimation of KL divergence. We vary the value of K and C, and find that K changes the absolute scale of KL divergence but does not affect the rankings of four conditions. Rankings among four conditions are consistent when C is small. We report results of $K = 1000, C = 10$ here. No stemming algorithm is performed and no stop words are removed, but case is ignored in the indexing process.

There are two kinds of variance in the estimation of divergence between two posterior distributions that should be carefully checked. The first kind of variance is attributed to Monte Carlo methods. We assess the Monte Carlo variance by calculating a 100α percent confidence interval as follows:

$$[\hat{D} - \Phi^{-1}(\frac{\alpha}{2})\frac{\hat{\sigma}}{\sqrt{M}}, \hat{D} + \Phi^{-1}(1 - \frac{\alpha}{2})\frac{\hat{\sigma}}{\sqrt{M}}],$$

where $\hat{\sigma}^2$ is the sample variance of $\theta_1, \theta_2, \ldots, \theta_M$, and $\Phi(\cdot)^{-1}$ is the inverse standard normal cumulative density function. The estimates of divergence using Monte Carlo integration is normally distributed due to the central limit theorem (Wasserman, 2004).

The second kind of variance is odue to the intrinsic uncertainties of data generating processes. We assess the second kind of variance by repeating each document collection pair with 1000 bootstrapped samples, i.e., sampling with replacement.

5.1.2.1 Quality of Monte Carlo Estimates

The Monte Carlo estimates of the KL divergence from several document collection pairs are listed in Table 5.1. We can see that the 95% confidence intervals capture well the Monte Carlo estimates of KL divergence. Other document collection pairs show similar tight confidence intervals – in fact, arbitrary, precision can be obtained by increasing the number of samples, M – and thus a complete list of the results is omitted. Because the Monte Carlo estimates are close to true values, we assume them to be exact and do not report the confidence intervals for the rest of the thesis.

$\mathcal{A}$	$\mathcal{B}$	$\hat{D}$	95% CI
ACQ	ACQ	2.76	[2.62, 2.89]
Palestinian	Palestinian	3.00	[3.54, 3.85]
Palestinian	Israeli	27.11	[26.64, 27.58]
Israeli	Palestinian	28.44	[27.97, 28.91]
Kerry	Bush	58.93	[58.22, 59.64]
ACQ	EARN	615.75	[610.85, 620.65]

Table 5.1: The Monte Carlo estimate $\hat{D}$ and 95% confidence interval (CI) of the Kullback-Leibler divergence of some document collection pairs $(\mathcal{A}, \mathcal{B})$ with the number of Monte Carlo samples $M = 1000$. The first row is Same Topic, the second row is Same Perspective, the third, fourth, and fifth rows are Different Perspectives, and the the sixth row is Different Topic.

Note that KL divergence is not symmetric. For example, the value of KL divergence of the pair (Israeli, Palestinian) is not necessarily the same as (Palestinian, Israeli). KL divergence is guaranteed to be greater than zero (Cover & Thomas, 1991) and is equal to zero only when document collections $\mathcal{A}$ and $\mathcal{B}$ are exactly the same. (ACQ, ACQ) is close to but not exactly zero because they are different samples of documents in the ACQ category.

5.1.2.2 Test of Different Ideological Perspectives

Now we present the main result of using distribution divergence to test if two document collections are written or spoken from difference perspectives. We calculate the KL divergence between posterior distributions of document collection pairs in four conditions using Monte Carlo methods, and plot the results in Figure 5.1.

The test of different perspectives based on statistical distribution divergence is shown to be very acute. The KL divergence of the document collection pairs in the DP condition fall mostly in the middle range, and is well separated from the high KL divergence of the pairs in DT condition and from the low KL divergence of the pairs in SP and ST conditions. By simply calculating the KL divergence of a document collection pair, we can predict that they were written from different perspectives if the value of KL divergence falls in the

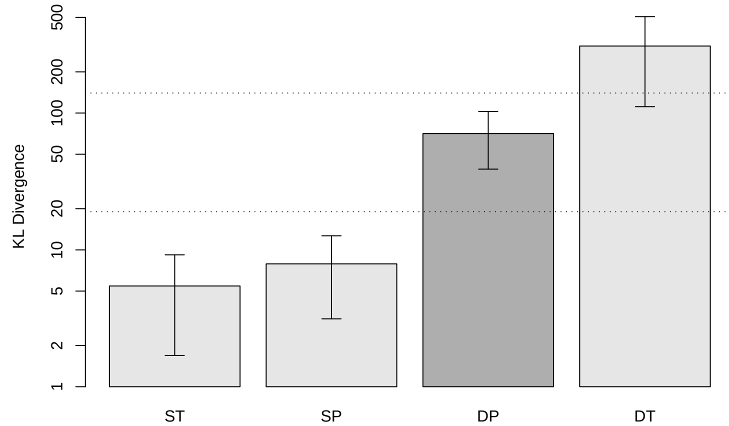


Figure 5.1: The values of KL divergence of the document collection pairs in four conditions: Different Perspectives (DP), Same Perspective (SP), Different Topics (DT), and Same Topic (ST). Note that the y axis is in log scale. The horizontal lines are drawn at the points with equivalent densities (based on Kernel Density Estimation).

middle range, from different topics if the value is very large, and from the same topic or perspective if the value is very small.

5.1.3 Personal Writing Styles or Ideological Perspectives?

One may suspect that the mid-range distribution divergence is attributed to personal speaking or writing styles and has nothing to do with different perspectives. The doubt is reasonable since half of the bitterlemons corpus was written by one Palestinian editor and one Israeli editor (see Table 3.2), and the debate transcripts were from only two candidates.

We test the theory and show that mid-range distribution divergence is indeed attributed to different perspectives by providing a counterexample, the document collection pair (Israeli Guest, Palestinian Guest) in the Different Perspectives condition. There are more than 200 different authors in each Israeli Guest or Palestinian Guest collection. How the distribution divergence of the pair is distributed can be mostly attributed to different perspectives, but cannot be attributed to writing styles. We compare the distribution divergence of the pair (Israeli Guest, Palestinian Guest) with others in Figure 5.2.

The results show that the distribution divergence of the (Israeli Guest, Palestinian

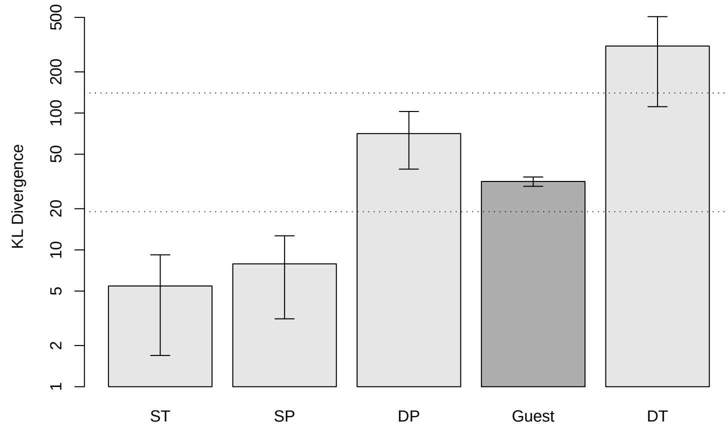


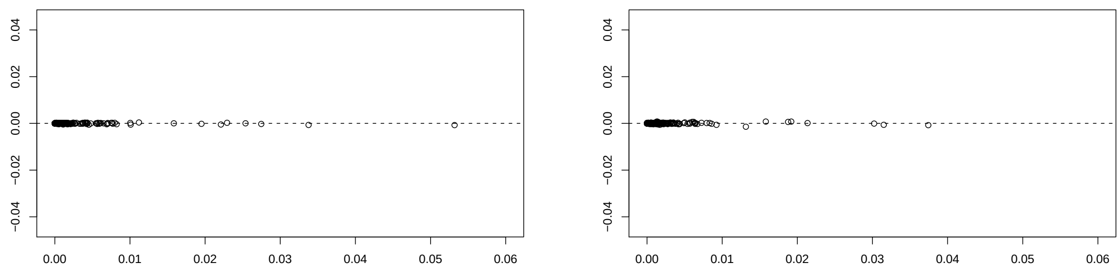
Figure 5.2: The average KL divergence of document collection pairs in the bitterlemons Guest subset (Israeli Guest vs. Palestinian Guest), ST, SP, DP, DT conditions. The horizontal lines are the same ones estimated in Figure 5.1.

Guest) pair, as other pairs in the DP condition, falls in the middle range, and is well separated from SP and ST in the low range and DT in the high range. Therefore, we are more confident that the test of different perspectives based on distribution divergence captures different perspectives, not personal writing or speaking styles.

5.1.4 Origins of the Differences

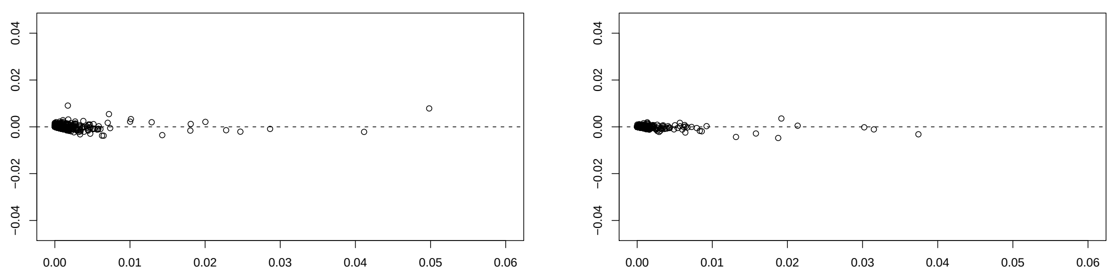
The effectiveness of the test of different perspectives is clearly demonstrated in Figure 5.1. However, one may wonder why the distribution divergence of the document collection pair with different perspectives falls in the middle range and what causes the large and small divergences of the document collection pairs with different topics (DT) and the same topic (ST) or perspective (SP), respectively.

We answer the question by taking a closer look at the causes of the distribution divergence in our model. We compare the expected marginal difference of θ between two posterior distributions $p(\theta|\mathcal{A})$ and $p(\theta|\mathcal{B})$. The marginal distribution of the i-th coordinate of θ, that is, the i-th word in the vocabulary, is a Beta distribution, and thus the expected value can be easily obtained. We plot the $\Delta\theta = \mathrm{E}[\theta_i|\mathcal{A}] - \mathrm{E}[\theta_i|\mathcal{B}]$ against $\mathrm{E}[\theta_i|\mathcal{A}]$ for each condition, and show samples in Figure 5.3.

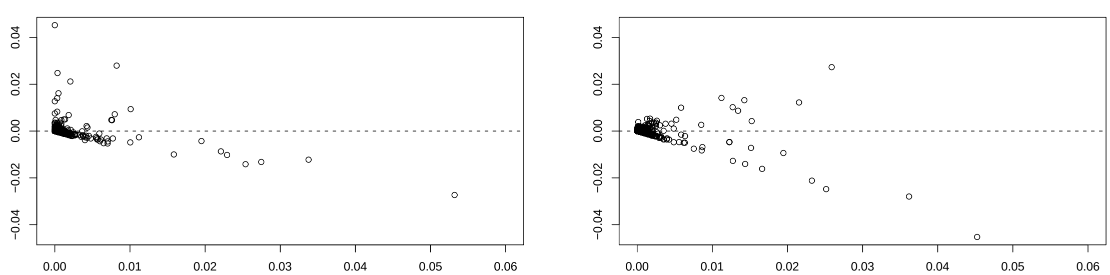


(a) Same Topic (ST)

(b) Same Perspective (SP)



(c) Different Perspective (DP)



(d) Different Topics (DT)

Figure 5.3: The $\Delta\theta$ (the y axis) v.s. θ (the x axis) plots of the typical document collection pairs in four conditions. The horizontal line is $\Delta\theta = 0$.

How $\Delta\theta$ deviates from zero not only reaffirms the unique statistical regularities of document collections of different perspectives, but also explains the origin of distribution divergence in other conditions. In Figure 5.3d we can see that the $\Delta\theta$ increases with the value of θ, and the deviance from zero is much greater than those in the Same Perspective (Figure 5.3b) and Same Topic (Figure 5.3a) conditions. The large $\Delta\theta$ not only accounts for large distribution divergence of the document pairs in DT conditions, but also marks distinct word distributions of document collections of different topics: a word that is frequent in one topic is less likely to be frequent in the other topic, and make readers perceive the topic of documents. At the other extreme, document collection pairs of the Same Perspective or Same Topic show very little difference in θ, which matches our intuition that documents of the same perspective or the same topic are similar.

The manner in which $\Delta\theta$ is varied with the value of θ in the Different Perspective condition is unique. The $\Delta\theta$ in Figure 5.3c is not as small as those in the SP and ST conditions, but at the same time not as large as those in DT conditions, resulting in mid-range distribution divergence in Figure 5.1. Why do document collections of different perspectives distribute this way? Since documents of different perspectives focus on a closely related issue (e.g., the Palestinian-Israeli conflict in the bitterlemons corpus, or the political and economical issues in the 2004 Presidential Debates corpus), they are expected to share a common vocabulary, but the different perspectives manifest themselves in subtle but not the same word emphasis. We list the top frequent words and their expected multinomial θ from Palestinian and Israeli documents in Table 5.2. We can see that the vocabulary from two perspectives highly overlaps with subtle difference in θ.

$\mathrm{E}[\theta\|\text{Palestinian}]$	$\mathrm{E}[\theta\|\text{Israeli}]$
palestinian (0.0394)	israel (0.0341)
israel (0.0372)	palestinian (0.0255)
state (0.0095)	state (0.0089)
politics (0.0077)	settle (0.0072)
peace (0.0071)	sharon (0.0071)
international (0.0066)	peace (0.0064)
people (0.0060)	arafat (0.0059)
settle (0.0057)	arab (0.0057)
occupation (0.0055)	politics (0.0051)
sharon (0.0055)	two (0.0050)
right (0.0054)	process (0.0044)
govern (0.0049)	secure (0.0043)
two (0.0047)	conflict (0.0039)
secure (0.0044)	lead (0.0039)
end (0.0042)	america (0.0035)
conflict (0.0042)	agree (0.0034)
process (0.0042)	right (0.0034)
side (0.0038)	gaza (0.0034)
negotiate (0.0038)	govern (0.0033)

Table 5.2: The statistical regularities of perspectives in text are highly overlapping vocabulary with subtle differences in frequencies.

5.2 Differentiating Ideological Video

We aim to develop a computer system that can automatically identify highly biased television news. Such system may increase audience's awareness about individual news broadcaster's bias and encourage them to seek news stories from contrasting viewpoints. However, it is not very clear at all that computers can automatically understand ideological perspectives expressed in television news.

- In this section, we develop a method of identifying differing ideological perspectives in news video based on what were chosen to be shown on screen. We motivate our method based on visual concepts in Section 5.2.1. We describe how to represent a video in terms of visual concepts (e.g., outdoor, car, and people walking) in Section 5.2.2.1, and how to quantify the similarity between two news video footages in terms of visual concepts in Section 5.2.2. A text corpora consists of words; here a television video consists of visual concepts.

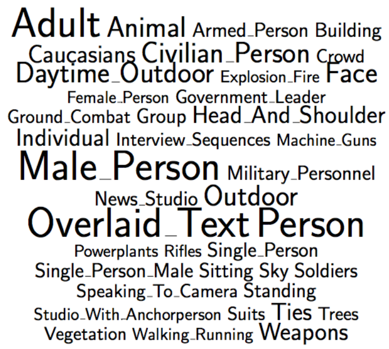
- We evaluate our method on a large broadcast news video archive (Section 3.2). To determine if two videos portray the same news event from differing ideological perspectives, we train a classifier to make a binary decision (i.e., same perspective or different perspectives). The classifier is shown to achieve high accuracy in Section 5.2.3.2. We apply the same idea to determine if two videos are about the same news event in Section 5.2.3.1.

- In Section 5.2.3.3, we conduct experiments to test if the high classification accuracy is indeed due to differences in choosing visual concept or broadcasters' idiosyncratic production styles.

- So far, our experiments are conducted using manual concept annotation to avoid a confounding factor arising from concept classifiers' poor performance. In Section 5.2.3.4, we repeat the above experiments and replace manual annotations with empirically trained concept classifiers.

5.2.1 Motivation

We are inspired by the recent work on developing large-scale concept ontology for video retrieval (A. G. Hauptmann, 2004) and consider a specific kind of visual grammar: *composition* (Efron, 1972). Research in media studies and communication provides us with some directions. News video footage, like paintings, advertisements, and illustrations, is not randomly designed and has its own visual "grammar" (Efron, 1972). Visual concepts



(a) CNN

(b) LBC

Figure 5.4: The text clouds show the frequencies of the visual concepts that were chosen by two broadcasters in the Iraq War stories. The larger a visual concept, the more frequently the concept was shown in news footage.

are generic objects, scenes, and activities (e.g., outdoor, car, and people walking). Visual concepts represent a video's visual content more closely than conventional low-level features (e.g., color, texture, and shape) can. Many researchers have actively developed concept classifiers to automatically detect concepts' presence in video. Therefore, if computers can automatically identify the visual concepts that are chosen to be shown in news footage, computers may be able to learn the difference between broadcasters' differing ideological perspectives.

We illustrate the idea in Figure 5.4. We count the visual concepts in the television news footage about the Iraq War from two different broadcasters (an American broadcaster CNN vs. an Arabic broadcaster LBC), and display them in text clouds (see Section 3.2 for more details about the data).

Due to the nature of broadcast news, it is not surprising to see many people-related visual concepts (e.g., Adult, Face, and Person). Because the news stories are about the Iraq War, it is also not surprising to see many war-related concepts (e.g., Weapons, Military Personnel, and Daytime Outdoor). The surprising differences, however, lie in the subtle emphasis on some concepts. Weapons and Machine Guns are shown more often in CNN (relative to other visual concepts in CNN) than in LBC. In contrast, Civilian Person and Crowd are shown more often in LBC than in CNN. How frequently some visual concepts are chosen seems to reflect a broadcaster's ideological perspective on a particular news event.

We thus hypothesize that news broadcasters holding different ideological beliefs choose to emphasize and de-emphasize some visual concepts when they portray a news event. We also hypothesize that each news broadcaster is self-consistent and chooses similar visual concepts to depict the same news event. Therefore, given two news videos, computers can compare the composition of news footage in terms of visual concepts. If two videos' composition is similar, they are probably produced by the same broadcaster (i.e., the same ideological perspective). However, if two videos' composition is not similar, they are probably produced by different news broadcasters (i.e., differing ideological perspectives). We formalize the idea and develop a computational method in Section 5.2.2.

The proposed method of identifying television videos conveying different ideological perspectives in this section is similar to that of identifying text corpora conveying different ideological perspectives in Section 5.1. We measure the "distance" between two text corpora, each of which contains multiple text documents from the same source, and use the distance as a classification feature. Similarly, for television videos, we measure the "distance" between two videos, each of which contain multiple shots, and use the distance as a classification feature. In text documents, the basic unit is a word; in television videos, the basic unit is a visual concept.

5.2.2 Measuring Semantic Similarity in Visual Content

To develop computer programs that can identify videos that convey differing ideological perspectives on a news event, we need to address the following two questions:

1. Can computers determine if two television news stories are about the same news event?

2. Given two television news stories on the same news event, can computers determine if they portray the event from differing ideological perspectives?

Although we could identify a news story's topic using textual clues (e.g., automatic speech recognition transcripts), we attack a more challenging question: grouping television news stories on the same event using only *visual* clues. More and more videos are produced and consumed by users on the Internet. Contrary to news videos, these web videos do not usually come with clear voice-over that describes what the video is about. An imagery-based topic tracking approach is more likely to be applicable for web videos than a text-based approach.

The two research questions can be boiled down to a single question:

> How well can we measure the *similarity* in visual content between two television news videos?

News videos on the same news event are likely to have *similar* visual content, while news videos on different news events are less likely to have *similar* visual content. Similarly, given two news videos on the same news event, broadcasters holding similar ideological beliefs are likely to portray the new event in a *similar* manner, while news broadcasters holding different ideological views are less likely to display *similar* visual content. Therefore, the key research question becomes how to measure the "semantic" *similarity* in visual content.

5.2.2.1 Representing Video as Visual Concepts

Our method measures semantic similarity between two news stories using a large-scale visual concept ontology. Our method consists of four steps, as illustrated in Figure 5.5. In Step 1, we first run a shot detector to detect shot boundaries in a news story, and select the middle frame of a shot as a key frame. In Step 2, we check if any concept in the visual concept ontology is present in the key frames. A concept's presence can be manually coded by human annotators, but it can also be coded using statistically trained concept classifiers. An example key frame and its visual concepts are shown in Figure 3.1.

We choose to represent the visual content of a television news story as a set of visual concepts shown on screen. By visual concepts, we mean generic objects, scenes, and activities (e.g., outdoor, car, and people walking). Low-level features (e.g., color, texture, shape) are easy to compute but fail to represent a video's visual content. For example, to compare how different broadcasters portray the Iraq War, knowing how many "soldiers" (a visual concept) they choose to show is much more informative than knowing how many brown patches (a low-level color feature) are shown.

We choose the Large-Scale Concept Ontology for Multimedia (LSCOM) (Kennedy & Hauptmann, 2006) to represent television video's visual content. LSCOM, initially developed for improving video retrieval, contains hundreds of generic activities, objects, and scenes[1]. The major categories and example concepts in each category are listed in

[1]The complete list of visual concepts is available at `http://www.lscom.org/concept.htm`

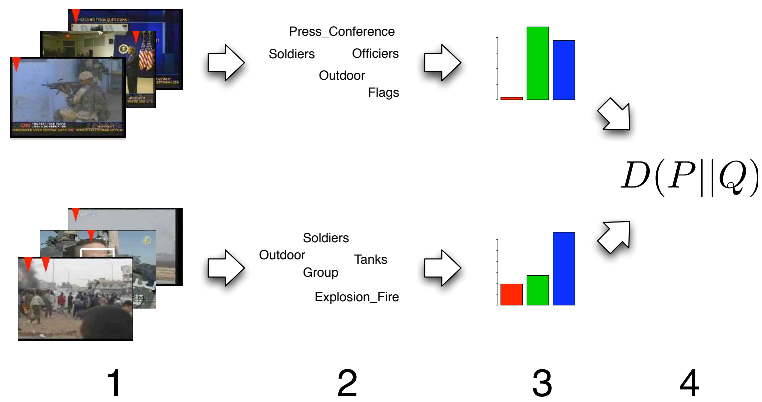


Figure 5.5: The method to measure semantic similarity in visual content consists of four steps. Step 1: extract the key frames of videos. Step 2: determine what visual concepts are present in key frames. Step 3: model visual concept occurrences using a multinomial distribution. Step 4: measure "distance" between two multinomial distributions using Kullback-Leibler divergence.

Table 5.2.2.1.

5.2.2.2 Measuring Similarity between Television News Videos

In Step 3 we model the occurrences of visual concepts in a news video's key frames using a statistical distribution. We take the same approach as we do for text corpora in Section 5.1.1. A text corpora consists of words; here a television video consists of visual concepts. We adopt bag-of-words representation for text corpora; here we adopt bag-of-concept representation for television news video. We model a text corpora as a multinomial distribution of words; here we model a television video as a multinomial distribution of visual concepts shown onscreen.

In Step 4, we measure the similarity between two videos' multinomial distributions in terms of Kullback-Leibler (KL) divergence (Cover & Thomas, 1991) as we do for measuring similarity between two text corpora in Section 5.1.1.

Category	Examples
Program	advertisement, baseball, weather news
Scene	indoors, outdoors, road, mountain
People	NBA players, officer, Pope
Objects	rabbit, car, airplane, bus, boat
Activities	walking, women dancing, cheering
Events	crash, explosion, gun shot
Graphics	weather map, NBA scores, schedule

Table 5.3: The major categories in LSCOM and sample concepts in each category.

5.2.3 Experiments on Differentiating Ideological Video

5.2.3.1 Identifying News Videos on the Same News Event

Because we are interested in how the *same* news event is portrayed by different broadcasters, we need to find the television news stories on the same news event in a video archive. As we argued in Section 5.2.2, this task boils down to comparing similarity between two videos' visual content. News videos on the same news event are likely to show *similar* visual content. Given two news videos, we could measure their similarity in terms of visual concepts as proposed in Section 5.2.2. If the value of KL divergence between two news videos is small (i.e., similar), they are likely to be on the same event.

We developed a classification task to evaluate our method of identifying news videos on the same event. There were two mutually exclusive categories in the classification task: *Different News Events* (DNE) vs. *Same News Event* (SNE). DNE contains news video pairs that are from the same broadcaster but on different news events (e.g., two videos from CNN: one is about the "Iraq War" and the other is about "Powell's resignation"). SNE contains news video pairs from the same broadcaster and on the same news event (e.g., two videos from CCTV about the same event "Tel Aviv bomb"). The predictor for the classification task is the value of KL divergence between two videos. We trained a binary classifier to predict if a news video pair belongs to SNE or DNE.

Among all possible video pairs that satisfy the conditions of Different News Event (DNE) and Same News Event (SNE), we randomly sampled 1000 video pairs for each category. We looked up their LSCOM concept annotations (Section 5.2.2.1), estimated multinomial distributions' parameters, and trained classifiers based on the values of (symmetric) KL divergence (see Section 5.2.2). We varied the training data size from 10% to 90%, and measured accuracy on the held-out 10% of video pairs. Accuracy is defined as

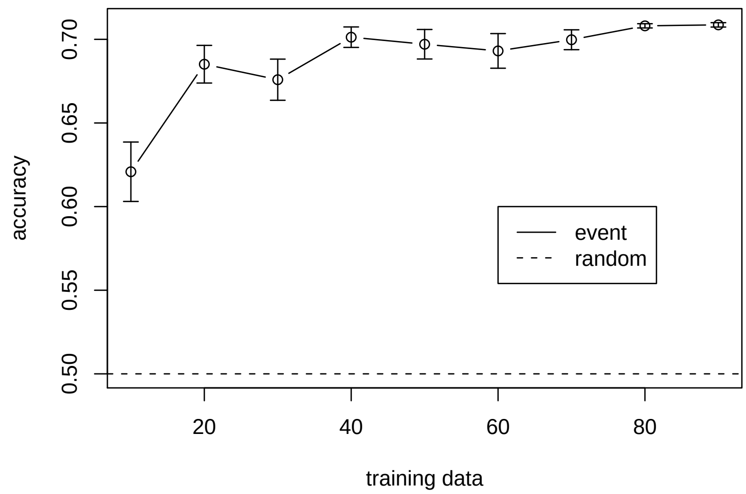


Figure 5.6: Our method can differentiate news video pairs on the same news event from the news video pairs on different news events significantly better than a random baseline. The x axis is the percentage of training data, and the y axis is the binary classification accuracy.

the number of video pairs that are correctly classified divided by the total number of video pairs in the held-out set. Because there were the same number of video pairs in each category, a random guessing baseline would have 50% accuracy. We repeated the experiments 100 times by sampling different video pairs, and calculated the average accuracy. The choice of classifier did not change the results much, and we reported only the results using Linear Discriminant Analysis and omitted the results using Support Vector Machines.

The experimental results in Figure 5.6 show that our method based on visual concepts can effectively distinguish news videos on the same news event from news videos on different news events. The classification accuracy was significantly better than the random baseline (t-test, $p < 0.01$) and reached a plateau around 70%. Our concept-based method of identifying television news stories on the same event could thus well complement other methods based on text (Allan, 2002), color (Zhai & Shah, 2005), semantic concepts (Zhang et al., 2004), and near-duplicates images (X. Wu et al., 2007). Although LSCOM was initially developed for supporting video retrieval, the results also suggest that LSCOM contain large and rich enough concepts to differentiate news videos on a variety of news events.

5.2.3.2 Identifying News Videos of Differing Ideological Perspectives

Given two news videos on the same news event, it is not clear if computers can tell whether they portray the event from different ideological perspectives. As we hypothesized in Section 5.2.1, given a news event, broadcasters holding similar ideological beliefs (i.e., the same broadcaster) are likely to choose similar visual concepts to compose news footage, while broadcasters holding different ideological beliefs (i.e., different broadcasters) are likely to choose different visual concepts. The task of identifying if two news videos convey differing ideological perspectives boils down to measuring if two videos are similar in terms of visual concepts (Section 5.2.2). If the value of KL divergence between two news videos about the same event is large (i.e., dissimilar), they are likely to come from broadcasters holding differing ideological beliefs.

We developed a classification task to evaluate our method of identifying news videos from differing ideological perspectives. There were two mutually exclusive categories in the classification task: *Different Ideological Perspectives* (DIP) vs. *Same Ideological Perspective* (SIP). DIP contains news video pairs that are about the same news event and from different broadcasters (e.g., two videos about "Arafat's death": one from LBC and one from NBC). SIP contains news video pairs that are about the same event but from the same broadcaster (e.g., two videos both from NTDTV about "Powell's resignation"). We trained a binary classifier to predict if a news video pair belongs to DIP or SIP. We followed the classification training and testing procedure in Section 5.2.3.1.

The experimental results in Figure 5.7 showed that our method based on visual concepts can effectively distinguish news videos produced by broadcasters with similar ideological beliefs from those with differing ideological beliefs. The classification accuracy was significantly better than a random baseline (t-test, $p < 0.01$), and reached a plateau around 72%. Given two news videos are on the same news event, we can then use our method to test if they portray the news from differing ideological perspectives.

We implicitly equate American news broadcasters with "American" ideological beliefs, and similarly Arabic news broadcaster with "Arabic" ideological beliefs. The assumptions follow from a particular choice of theory of ideology (see Section 1.1). Different definitions of ideology may result in different interpretations of our experiment results such as distinguishing national news broadcasters.

Although our method achieved significant improvement over the random baseline, there was considerable room for improvement. We focused on the visual concepts chosen differently by individual news broadcasters, but this did not exclude possibilities for improving the classification by incorporating signals other than visual concepts. For example, broadcast news videos contain spoken words from anchors, reporters, or interviewees, and the word choices have been shown to exhibit a broadcaster's ideological perspectives

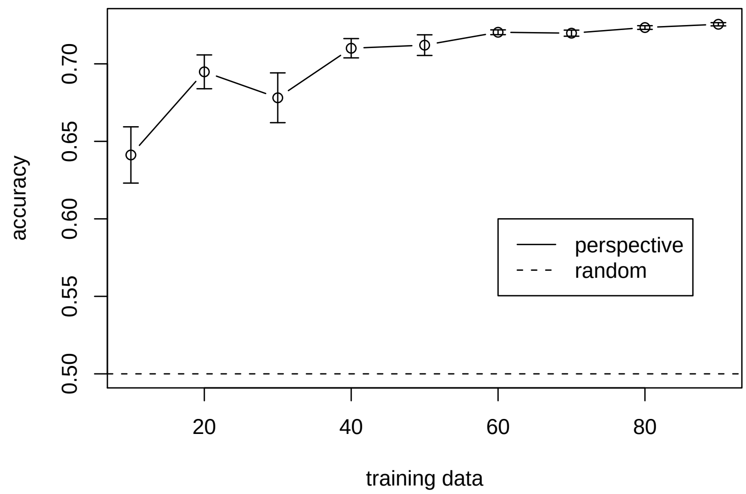


Figure 5.7: Our method can differentiate the news video pairs conveying different ideological perspectives from the news videos conveying similar ideological perspectives significantly better than a random baseline. The x axis is the percentage of training data, and the y axis is the binary classification accuracy.

(Lin & Hauptmann, 2006).

Because we already knew a video's broadcaster when the video was recorded, was the task of determining whether two news videos portray the news event from differing ideological perspectives as trivial as checking if they came from different broadcasters?

- Although we can accomplish the same task using metadata such as a news video's broadcaster, this method is unlikely to be applicable to videos that contain little metadata (e.g., web videos on YouTube). We opted for a method of broader generalization and developed our method solely based on visual content and generic visual concepts.

- We chose the TRECVID'05 video achieve because the ideological perspective from which the news videos were produced (i.e., its broadcaster) were clearly labeled. The clearly labeled data allowed us to conduct controlled experiments.

5.2.3.3 Broadcasters' Idiosyncratic Production Styles?

The experimental results in Section 5.2.3.2 seem to suggest that broadcasters with differing ideological beliefs choose different imageries to portray the same news event. However, we can provide an alternative theory for the high classification accuracy. Each broadcaster usually has idiosyncratic production styles (e.g., adding station logo in the corner, unique studio scenes, etc.) and a fixed number of anchors and reporters. However, news video pairs in the DIP category in Section 5.2.3.3 could result in large values of KL divergence because broadcasters' production styles are very different from each other, while the news videos in the SIP category could result in small values of KL divergence because the same broadcaster shares the similar production artifacts. Is it possible that the classifiers in Section 5.2.3.2 learned only broadcasters' idiosyncratic production styles to determine if they portray a news event differently?

We developed the following classification task to test the theory. There were two mutually exclusive categories: *Different Events and Different Ideological Perspectives* (DEDIP) vs. *Different Events and Similar ideological Perspective* (DESIP). The two categories were similar to the DIP vs. SIP contrast in Section 5.2.3.2, and the only difference was that the DIP vs. SIP contrast contained video pairs about the same news event, while the DEDIP vs. DESIP contrast in this section contained video pairs covering different news events. If the theory of broadcasters' idiosyncratic production styles holds true for the DIP vs. SIP contrast, it should also hold true for the DEDIP and DESIP contrast. Therefore, we would expect the classification accuracy to be as high as the accuracy in Section 5.2.3.2. We followed the classification training and testing procedure in Section 5.2.3.2.

The experimental results in Figure 5.8 show that it is very unlikely that the high classification accuracy in Section 5.2.3.2 is due to broadcasters' idiosyncratic production styles. The classification accuracy is slightly better than a random baseline (t-test, $p < 0.02$) but very close to random. The production styles seem to contribute to classifying whether or not news video pairs come from the same broadcasters, but the magnitude was minor and cannot fully account for the high accuracy achieved in Section 5.2.3.2.

5.2.3.4 Concept Classifiers' Accuracy

So far our experiments were based on manual annotations of visual concepts from LSCOM. Using manual annotation is equivalent to assuming that perfect concept classifiers are available. The state-of-the-art classifiers are far from perfect for most visual concepts (M. R. Naphade & Smith, 2004). So how well can computers determine if two news videos convey a differently ideological perspective on a news event using empirically trained classifiers?

We obtained empirical accuracies of 449 LSCOM concept classifiers by training Sup-

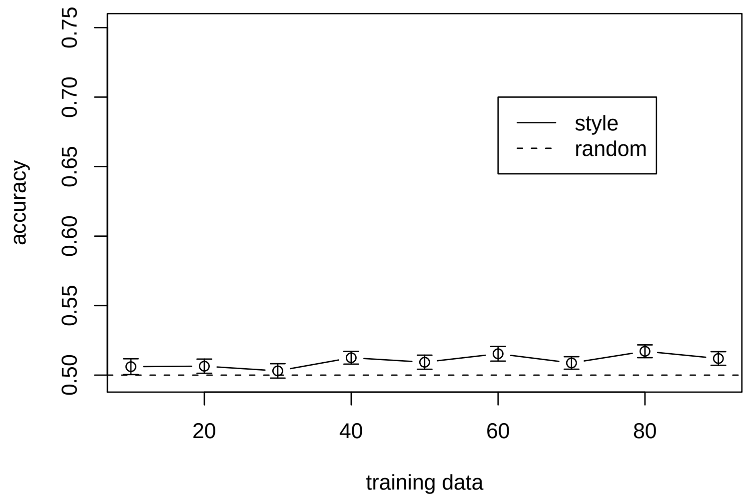
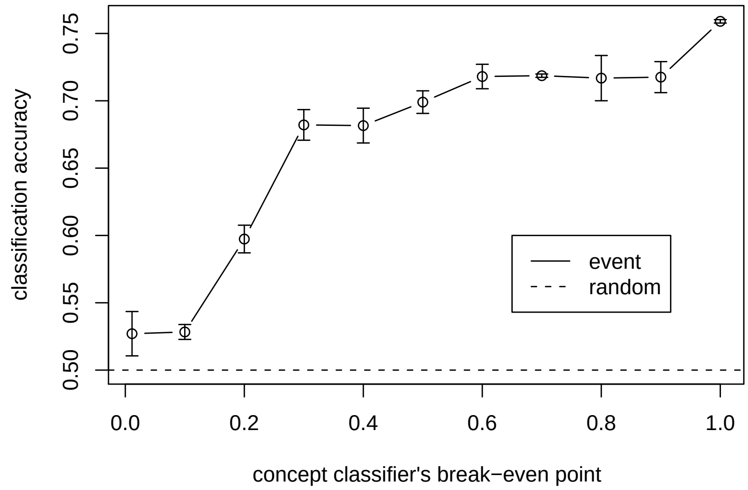


Figure 5.8: The contrast between DEDIP and DESIP did not achieve as high accuracy as that in Section 5.2.3.2. The x axis is the percentage of training data, and the y axis is the binary classification accuracy.

port Vector Machines on 90% of positive examples and testing on the held-out 10% of examples. We first trained uni-modal concept classifiers using single low-level features (e.g., color histogram in various grid sizes and color spaces, texture, text, audio, etc.). We then built multi-modal classifiers that fused the outputs from best uni-modal classifiers (see (A. G. Hauptmann et al., 2005) for more details about the training procedure). We evaluated the performance of the best multi-modal classifiers on the held-out set in terms of average precisions (AP).

We varied the concept classifier's accuracy by injecting noise into manual annotations. AP is a rank-based evaluation metric, but our experiments relied on set-based metrics. We thus approximated AP using recall-precision break-even points (or R-precision), which was highly correlated with AP (Manning, Raghavan, & Schütze, 2008). We randomly flipped the positive and negative labels of visual concepts until we reached the desired break-even points. We varied the classifier's break-even points ranging from APs obtained from empirically trained classifiers to 1.0, and repeated the experiments in Section 5.2.3.2 and Section 5.2.3.1.

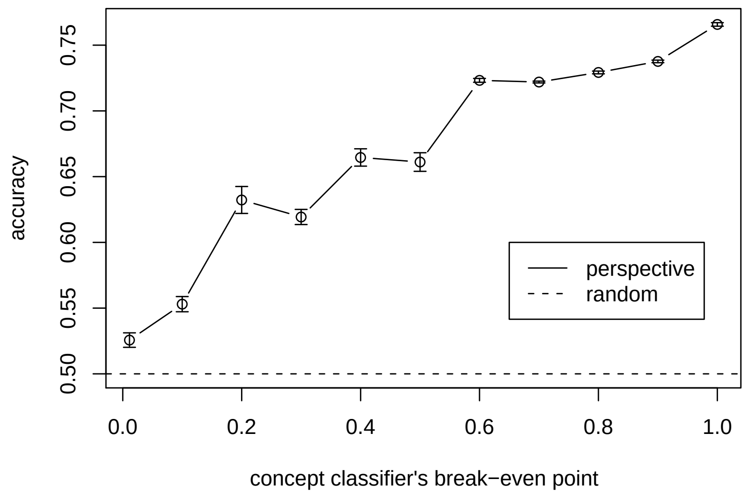
The experimental results showed that the empirically trained classifiers cannot satisfactorily identify news videos covering the same news event (Figure 5.9a) and news videos



(a) Identifying news video pairs covering similar news events. The accuracy of classifying video pairs into same event/different event categories (y axis) increases as the accuracy of the visual concept classifier improves (x axis). The performance at the accuracy 1.0 of visual concepts classifier on the x axis establishes the upper bound of the approach based on visual concepts.

Figure 5.9: We varied the classifier's accuracy and repeated the two experiments in Figure 5.6 and Figure 5.7. The x axis is the (simulated) classifier's accuracy in terms of precision-recall break-even points. The leftmost data point was based on the performance of the empirically trained classifiers. The y axis is the classification accuracy.

conveying different perspectives (Figure 5.9b). The median AP of the empirically trained classifiers was 0.0113 (i.e., the x coordinate of the leftmost data point in Figure 5.9). Although the classification accuracy using empirically trained concept classifiers (i.e., the leftmost data point) was statistically significant from random (t-test, $p < 0.01$), the accuracy was very close to random and unlikely to make a realistic difference. It was not surprising that the classification accuracy improved as concept classifier's break-even points increased. To achieve reasonable performance, we seemed to need a concept classifier of break-even points 0.6. We should not be easily discouraged by current classifiers' poor performance. With the advances in computational power and statistical learning algorithms, it is likely that the concept classifier's accuracy will be continuously improved. Moreover, we may be able to compensate for poor accuracy by increasing the number of



(b) Identifying news video pairs of different ideological perspectives. The accuracy of classifying video pairs into same viewpoint/different viewpoints categories (y axis) increases as the accuracy of the visual concept classifier improves (x axis). The performance at the accuracy 1.0 of visual concepts classifier on the x axis establishes the upper bound of the approach based on visual concepts.

concepts, as demonstrated recently in the study of improving video retrieval using more than thousands of visual concepts (A. Hauptmann, Yan, & Lin, 2007).

Chapter 6

Identifying Ideological Perspectives at the Document Level

Polarizing discussions on political and social issues are common in mass and user-generated media. However, computer-based understanding of ideological discourse has been considered too difficult to undertake. In this chapter, we present a statistical model for understanding ideological discourse. By ideology, we mean "a set of general beliefs socially shared by a group of people." For example, Democratic and Republican are two major political ideologies in the United States. Our model captures lexical variations due to an ideological text's topic and due to an author or speaker's ideological perspective. To cope with the non-conjugacy of the logistic-normal prior, we derive a variational inference algorithm for the model. We evaluate our model on synthetic data as well as written and spoken political discourses. Experimental results strongly support that ideological perspectives are reflected in lexical variations.

- In this chapter, we developed a statistical model for ideological discourse. Based on the empirical observation in Section 5.2.1, we hypothesized that ideological perspectives were reflected in lexical variations. Some words were used more frequently because they were highly related to an ideological text's topic (i.e., *topical*), while some words were used more frequently because authors holding a particular ideological perspective chose so (i.e., *ideological*).

- We formalized a hypothesis and developed a statistical model for ideological discourse in Chapter 4. Lexical variations in ideological discourse were encoded in a word's topical and ideological weights. The coupled weights and the non-conjugacy of the logistic-normal prior posed a challenging inference problem. We developed an approximate inference algorithm based on the variational method in Section 4.2.

- We evaluated our model on synthetic data (Section 4.5.1) as well as on written text and spoken text (Section 6.2.3.1). In Section 4.5.3, we showed that our model automatically uncovered many discourse structures in ideological discourses identified by researchers in discourse analysis (Van Dijk, 1998)

- In Section 4.5.4, we showed that our model fit ideological corpora better than a model that assumes no lexical variations due to an author or speaker's ideological perspective. Therefore the experimental results strongly suggested that ideological perspectives were reflected in lexical variations.

In Section 6.1, we investigate a problem of identifying the ideological perspective from which a document is written. By perspective, we mean a point of view, for example, from the perspective of Democrats or Republicans. Can computers learn to identify the perspective of a document?

We develop statistical models to capture how perspectives are expressed at the document level and evaluate our models on articles about the Israeli-Palestinian conflict. The results show that our models successfully learn how perspectives are reflected in word usage and can identify the perspective of a document with high accuracy.

Television news has become the predominant way of understanding the world, but individual news broadcasters can frame or mislead an audience's understanding of political and social issues. We are developing a computer system that can automatically identify highly biased television news and encourage audiences to seek news stories from contrasting viewpoints. But it is not very clear if computers can identify the ideological perspective from which a news video was produced. In Section 6.2, we present a method based on empathic patterns of visual concepts: news broadcasters with contrasting ideological beliefs appear to emphasize different subsets of visual concepts. We formalize the emphatic patterns and develop a statistical model. We evaluate our model on a large broadcast news video archive, and show that the experimental results are promising.

We develop a classifier that can automatically identify a web video's ideological perspective on a political or social issue (e.g., pro-life or pro-choice on the abortion issue). The problem has received little attention, possibly due to inherent difficulties in content-based approaches. In Section 6.3 we develop such a classifier based on the pattern of tags emerging from folksonomies. The experimental results are positive and encouraging.

6.1 Identifying Ideological Perspectives in Text

In this section, we investigate a new problem of automatically identifying the ideological perspective from which a document is written.

We approach the problem of learning individual perspectives in a statistical framework. We develop statistical models to learn how perspectives are reflected in word usage, and we treat the problem of identifying perspectives as a classification task. Although our corpus contains document-level perspective annotations, it lacks sentence-level annotations, creating a challenge for learning the perspective of sentences. We develop a novel statistical model to overcome this problem. The experimental results show that our statistical models can successfully identify the perspective from which a document is written with high accuracy.

6.1.1 Statistical Modeling of Perspectives

We develop algorithms for learning perspectives using a statistical framework. We denote a training corpus to be a set of documents W_n and their perspectives labels D_n, $n = 1, \ldots, N$, where N is the total number of documents in the corpus. Given a new document $\tilde{W}$ with unknown document perspective, the perspective $\tilde{D}$ is calculated based on the following conditional probability.

$$P(\tilde{D}|\tilde{W}, \{D_n, W_n\}_{n=1}^{N}) \tag{6.1}$$

We are also interested in how strongly each sentence in a document conveys perspective information. We denote the opacity of the m-th sentence of the n-th document as a binary random variable $S_{m,n}$. To evaluate $S_{m,n}$, how strongly a sentence reflects a particular perspective, we calculate the following conditional probability.

$$P(S_{m,n}|\{D_n, W_n\}_{n=1}^{N}) \tag{6.2}$$

We model the process of generating documents from a particular perspective as follows:

$$\begin{aligned} \pi &\sim \mathrm{Beta}(\alpha_\pi, \beta_\pi) \\ \theta &\sim \mathrm{Dirichlet}(\alpha_\theta) \\ D_n &\sim \mathrm{Binomial}(1, \pi) \\ W_n &\sim \mathrm{Multinomial}(L_n, \theta_d) \end{aligned}$$

First, the parameters π and θ are sampled once from prior distributions for the whole corpus. Beta and Dirichlet are chosen because they are conjugate priors for binomial and multinomial distributions, respectively. We set the hyper-parameters α_π, β_π, and α_θ to one, resulting in non-informative priors. A document perspective D_n is then sampled from a binomial distribution with the parameter π. The value of D_n is either d^0 (Israeli) or d^1 (Palestinian). Words in the document are then sampled from a multinomial distribution, where L_n is the length of the document. A graphical representation of the model is shown in Figure 6.1.

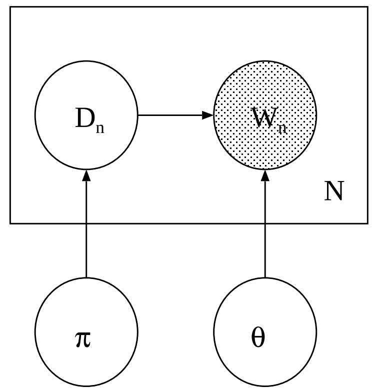


Figure 6.1: naïveBayes Model

The model described above is commonly known as a naïveBayes (NB) model. NB models have been widely used for various classification tasks, including text categorization (Lewis, 1998). We use the NB model as a building block for our later model that incorporates sentence-level perspective information.

To predict the perspective of an unseen document using naïveBayes, we calculate the posterior distribution of $\tilde{D}$ in (6.1) by integrating out the parameters,

$$\int\int P(\tilde{D}, \pi, \theta | \{(D_n, W_n)\}_{n=1}^{N}, \tilde{W}) d\pi d\theta \tag{6.3}$$

However, the above integral is difficult to compute. As an alternative, we use Markov Chain Monte Carlo (MCMC) methods to obtain samples from the posterior distribution. Details about MCMC methods can be found in Appendix A.

6.1.2 Experiments on Identifying Ideological Perspectives in Text

We evaluate three different models for the task of identifying perspectives at the document level: two naïveBayes models (NB) with different inference methods, NB-B and NB-M, and a Support Vector Machines (SVM) (Cristianini & Shawe-Taylor, 2000). NB-B uses full Bayesian inference and NB-M uses Maximum a posteriori (MAP). We compare NB with SVM which has been very effective for classifying topical documents (Joachims, 1998), and contrast NB, generative models, with SVM, discriminative models. For training SVM, we represent each document as a V-dimensional feature vector, where V is the vocabulary size and each coordinate is the normalized term frequency within the document. We use a linear kernel for SVM and search for the best parameters using grid methods.

To evaluate the statistical models, we train them on the documents in the bitterlemons corpus and calculate how accurately each model predicts document perspective in ten-fold cross-validation experiments. Table 6.1 reports the average classification accuracy across the 10 folds for each model. The accuracy of a baseline classifier, which randomly assigns the perspective of a document as Palestinian or Israeli, is 0.5 because there are the same numbers of documents from the two perspectives.

SVM is a popular binary classifier that has been shown to be very effective in text classification (Joachims, 1998). Unlike generative probabilistic models like naïveBayes model, SVM is based on the idea of finding a hyperplane in the feature space that separates two classes of data points while keeping the margin as large as possible. More details about SVM can be found in (Cristianini & Shawe-Taylor, 2000). To train SVM we represent each document as a V-dimensional feature vector, where V is the vocabulary size and each coordinate is the normalized term frequency. We use linear kernel for SVM and find the best parameters using grid search. Our SVM implementation is based on `LIBSVM` (Chang & Lin, 2001).

Model	Data Set	Accuracy	Reduction
Baseline		0.5	
SVM	Editors	0.9724	
NB-M	Editors	0.9895	61%
NB-B	Editors	0.9909	67%
SVM	Guests	0.8621	
NB-M	Guests	0.8789	12%
NB-B	Guests	0.8859	17%

Table 6.1: Results for Identifying Perspectives at the Document Level

The last column of Table 6.1 is error reduction relative to SVM. The results show that the naïveBayes model and SVM perform surprisingly well on both the Editors and Guests subsets of the bitterlemons corpus. The naïveBayes models make fewer errors than SVM, possibly due to properties of achieving optimality with the size of a training set between discriminative models and generative models (A. Y. Ng & Jordan, 2002). NB-B, which performs full Bayesian inference, improves on NB-M, which only performs point estimation. The results suggest that the choice of words made by the authors, either consciously or subconsciously, reflects much of their political perspectives. Statistical models can capture word usage well and can identify the perspective of documents with high accuracy.

Given the performance gap between Editors and Guests subsets, one may argue that there exist distinct editing artifacts or writing styles between editors and guests, and that the statistical models capture those artifacts or writing styles rather than “perspectives.” To test if the statistical models are truly learning perspectives, we conduct experiments in which the training and testing data are mismatched, i.e., from different subsets of the corpus. If what the SVM and naïveBayes model learn are writing styles or editing artifacts, the classification performance under the mismatched conditions will be considerably degraded.

Model	Training	Testing	Accuracy	
Baseline			0.5	
SVM	Guests	Editors	0.8822	
NB-M	Guests	Editors	0.9327	43%
NB-B	Guests	Editors	0.9346	44%
SVM	Editors	Guests	0.8148	
NB-M	Editors	Guests	0.8485	18%
NB-B	Editors	Guests	0.8585	24%

Table 6.2: Identifying Document-Level Perspectives with Different Training and Testing Sets

The results on the mismatched training and testing experiments are shown in Table 6.2. Both SVM and the two variants of naïveBayes model perform well on the different combinations of training and testing data. As in Table 6.1, the naïveBayes models perform better than SVM with larger error reductions, and NB-B slightly outperforms NB-M. The high accuracy on the mismatched experiments suggests that statistical models do not learn writing styles or editing artifacts. This suggests that the perspective of a documents is mostly reflected in how words are chosen by the writers.

We list the most frequent words (excluding stop words) learned by the naïveBayes model in Table 6.3. The frequent words overlap greatly between the Palestinian and Israeli perspectives, including “state,” “peace,” “process,” “secure” (“security”), and “govern” (“government”). This is in contrast to what we expect from topical text classification (e.g., “Sports” vs. “Politics”), in which frequent words seldom overlap. Authors from different perspectives often choose words from a similar vocabulary but emphasize them differently. For example, in documents that are written from the Palestinian perspective, the word “palestinian” is mentioned more frequently than the word “israel.” It is, however,

Palestinian	palestinian, israel, state, politics, peace, international, people, settle, occupation, sharon, right, govern, two, secure, end, conflict, process, side, negotiate
Israeli	israel, palestinian, state, settle, sharon, peace, arafat, arab, politics, two, process, secure, conflict, lead, america, agree, right, gaza, govern

Table 6.3: The top twenty most frequent stems learned by naïveBayes model, sorted by $P(w|d)$

the opposite for documents that are written from the Israeli perspective. Perspectives are also expressed by how frequently certain people ("sharon" vs. "arafat"), countries ("international" vs. "america"), and actions ("occupation" vs. "settle") are mentioned. While one might solicit these contrasting word pairs from domain experts, our results show that statistical models such as SVM and naïveBayes model can automatically acquire them.

6.2 Identifying Ideological Perspectives in Television News

We develop a computer system that can automatically identify highly biased television news. Such a system may increase the audience's awareness about individual news broadcaster's bias, and can encourage them to seek news stories from contrasting viewpoints. Considering multiple viewpoints could help people make more informed decisions and strengthen democracy. However, it is not clear if computers can automatically understand the ideological perspectives expressed in television news.

- In this section, we present a method of automatically identifying the ideological perspective (e.g., American vs. non-American) from which a news video about a particular news event (e.g., the Iraq War) is produced. Our method is based on a pattern described in Section 6.2.1: news broadcasters with contrasting ideological beliefs seem to emphasize different visual concepts. By visual concepts we mean generic scenes, objects, and actions (e.g., Outdoor, Car, and People Walking).

- We formalize the emphatic patterns of visual concepts and develop a statistical model in Section 6.2.2. The model simultaneously model what concepts are shown more frequently because they are highly related to a news event and what concepts are emphasized or de-emphasized due to a news broadcaster's ideological perspective. Each visual concept is assigned a topical weight and ideological weights. The coupled weights, however, make statistical inference very difficult. We thus develop an approximate inference algorithm based on variational method in Section 4.2. We describe how to apply the model to predict the ideological perspective of an unidentified news video in Section 4.4.

- We evaluate our method on a large broadcast news video archive (Section 6.2.3.1) using binary classification tasks. Given a news topic, we train a perspective classifier (e.g., the American view on the Iraq War), and evaluate the classifier on a held-out set. We show that our model achieves high accuracy of predicting a news video's ideological perspective in Section 6.2.3.2. We also give examples of emphatic patterns of visual concepts in Section 6.2.3.3 that are automatically learned from a video news archive.

- The high perspective classification accuracy in Section 6.2.3.2, however, could be attributed to individual news broadcaster's production styles, and have little to do with ideological perspectives. In Section 6.2.3.4, we test the theory, and show that production styles, although they exist, cannot completely account for non-trivial perspective classification accuracy.

- So far we have conducted the experiments using manual visual concept annotations to avoid a confounding factor arising from the concept classifiers' poor performance. In Section 6.2.3.5, we relax the assumption and repeat the above experiments using empirically trained concept classifiers.

6.2.1 Emphatic Patterns of Visual Concepts

News video footages, like paintings, advertisements, and technical illustrations, are not randomly put together, but have their own visual "grammar" (Kress & Leeuwen, 1996). What rules of the visual grammar in news video production can we use to identify the ideological perspective from which a news video was produced? We seek a visual grammar rule that not only discriminates between one ideology and another, but also can be automatically computed without much human intervention.

We consider a specific visual grammar rule in news video production: *composition* (Kress & Leeuwen, 1996). Composition rules define what entities are chosen to compose a visual display. Inspired by the recent work on developing a large-scale concept ontology for video retrieval (A. G. Hauptmann, 2004), we characterize entities in terms of "visual concepts." Visual concepts are generic objects, scenes, and activities (e.g., Outdoor, Car, and People Walking). Visual concepts characterize visual content more closely than low-level features (e.g., color, texture, and shape) can. Figure 3.1 shows a set of visual concepts chosen to be shown in a key frame.

Do news broadcasters with contrasting ideological beliefs differ in composing news footages about a particular news event, specifically, in terms of choosing visual concepts onscreen? We first present empirical observations on the news footages about the Iraq War. We count the visual concepts shown in the news footages from two groups of news broadcasters with contrasting ideological beliefs, and display them in text clouds in Figure 6.2 (American news broadcasters) and Figure 6.3 (Arabic news broadcasters). The larger a visual concept's size, the more frequently the concept was shown in news footage. See Section 6.2.3.1 for more details about the data.

Adult Armed_Person Building Caucasians Civilian_Person Crowd Daytime_Outdoor Face Female_Person Government_Leader Ground_Combat Group Head_And_Shoulder Individual Interview_Sequences Machine_-Guns Male_Person Military_Personnel Outdoor Overlaid_Text Person Powerplants Rifles Single_Person Single_Person_Male Sitting Sky Smoke Soldiers Speaking_To_Camera Standing Street_Battle Suits Talking Ties Trees Urban_Scenes Vegetation Walking_Running Weapons

Figure 6.2: The text cloud shows the frequency of the top 10 percent most frequent visual concepts that were chosen by *American* news broadcasters in the Iraq War news footage.

Adult Building Civilian_Person Crowd Daytime_Outdoor Face Female_Person Furniture Government_Leader Ground_Vehicles Group Head_And_Shoulder Individual Interview_On_Location Male_Person Meeting Microphones Military_Personnel Muslims Outdoor Person Politics Powerplants Press_Conference Road Scene_Text Single_Person Single_Person_Male Sitting Sky Standing Suits Talking Ties Urban_Scenes Vegetation Vehicle Weapons Windows

Figure 6.3: The text cloud shows the frequency of the top 10 percent of visual concepts that were chosen by *Arabic* news broadcasters in the Iraq War news footage.

Due to the nature of broadcast news, it is not surprising to see many people-related visual concepts (Adult, Face, and Person) in both American and Arabic news media. In addition, because the news stories are about the Iraq War, it is not surprising to see many war-related concepts (Weapons, Military_Personnel, and Daytime_Outdoor). What is most surprising is in the subtle "emphasis" on a subset of concepts. Weapons and Machine_Guns are chosen to be shown more often (i.e., large word size) in American news broadcasts (relative to other visual concepts in American news media) than in Arabic news broadcasts. In contrast, Civilian_Person and Crowd are chosen to be shown more often in Arabic news media than in American news media. How frequently certain visual concepts (Weapons vs. Civilian_Person) are emphasized seems to reflect a broadcaster's ideological perspective (American view vs. Arabic view) on a particular news event (the Iraq War).

We thus hypothesize that news broadcasters with different ideological beliefs emphasize (i.e., show more frequently) or de-emphasize (i.e., show less frequently) certain visual concepts when they portray a news event. Some visual concepts are shown more frequently because they are highly related to a specific news topic regardless of news broadcasters, and we call these concepts *topical* (e.g., Military_Personnel and Daytime_Outdoor for the Iraq War news). Some visual concepts are shown more frequently because news broadcasters with a particular ideological perspective choose so in portraying a particular news event (e.g., Weapons in American news media vs. Civilian_Person in Arabic news media for the Iraq War news), and we call these concepts *ideological*.

Many researchers have actively developed visual concept classifiers to automatically detect visual concepts' presence in image and video (M. R. Naphade & Smith, 2004). If computers can automatically identify many visual concepts that are chosen to be shown in news footage, computers may be able to learn the compositional difference between news broadcasters with contrasting ideological beliefs. By comparing the emphatic patterns of visual concepts in a news video, computers can automatically predict a news video's ideological perspective. We formalize emphatic patters of visual concepts in a statistical model in Section 6.2.2.

6.2.2 Joint Topic and Perspective Models for News Videos

We capture the emphatic patterns of visual concepts exhibited in ideological news videos using a statistical model described in Chapter 4. In Section 6.2.1, we identify two factors that make up the emphatic pattern of visual concepts: *topical* and *ideological*. In the statistical model, we assign a *topical* weight to each visual concept to represent how frequently the concept is chosen because of a news topic or event (e.g., Outdoor is shown more frequently in the Iraq War news). In this statistical model, we assign *ideological* weights to each visual concept to represent to what degree a visual concept is emphasized or de-emphasized by news broadcasters holding a particular ideological perspective (e.g., Weapon is shown more frequently in the American news for the Iraq War news stories). We aim to uncover these topical and ideological weights simultaneously from data. Furthermore, we apply the model and learned topical and ideological weights to predict the ideological perspective from which an unidentified news video is produced.

We illustrate the idea of the Joint Topic and Perspective Model in the context of television news in a three-visual-concept world in Figure 6.4. Any point in the three-visual-concept simplex represents the proportion of three visual concepts (i.e., Outdoor, Weapon, and Civilian) chosen to be shown in news footage (also known as a multinomial distribution's parameter). Let T denote the proportion of the three concepts for a particular news topic (e.g., the Iraq War). T represents how likely an audience would see

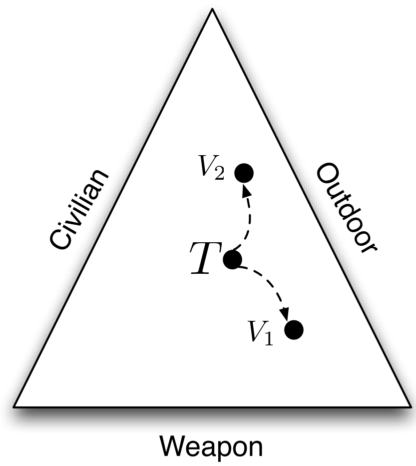


Figure 6.4: A three-visual-concept simplex illustrates the main idea behind the Joint Topic and Perspective Model for news videos. T denotes the proportion of the three concepts (i.e., topical weights) that are chosen to be shown on screen for a particular news topic. V_1 denotes the proportion of the three concepts after the topical weights are modulated by news broadcasters holding one particular ideological perspective; V_2 denotes the proportion of the weights modulated by news broadcasters holding the other particular set of ideological beliefs.

Outdoor, Weapon, or Civilian in the news footage about the Iraq War. Now suppose a group of news broadcasters holding a particular ideological perspective choose to show more Civilian and fewer Weapon. The *ideological* weights associated with this group of news broadcasters in effect move the proportion from T to V_1. When we sample visual concepts from a multinomial distribution of a parameter at V_1, we would see more Civilian and fewer Weapon. Now suppose a group of news broadcasters holding a contrasting ideological perspective choose to show more Weapon and fewer Civilian. The *ideological* weights associated with this second group of news broadcasters move the proportion from T to V_2. When we sample visual concepts from a multinomial distribution of a parameter at V_2, we would see more Weapon and fewer Civilian onscreen. The topical weights determine the position of T in a simplex, and each ideological perspective moves T to a biased position according to its ideological weights.

6.2.3 Experiments

6.2.3.1 News Videos and Visual Concepts

We evaluated our method of identifying news videos' ideological perspectives on a broadcast news video archive from the 2005 TREC Video Evaluation (TRECVID) (Over et al., 2005) (see Section 3.2.1).

We focused on news story footage, and removed non-news segments (e.g., music videos, drama, commercials, etc.) and news production-related scenes (e.g., anchors, news studio, etc.) based on LSCOM annotations. These non-news and news production related scenes were removed because they represented mostly broadcasters' production styles and conveyed very little about their ideological beliefs.

We categorized the news videos in the TRECVID 2005 archive into three ideological perspectives: American, Arabic, and Chinese. The news broadcasters and news channels in each ideology are listed in Table 6.4. Although a news broadcaster may express a distinct attitude on a news event, we did not consider the news broadcasters' personal ideologies. We are more interested in cultural, historical, and social beliefs and values that are broadly shared across news broadcasters.

Ideology	Hours	News Broadcaster (Channel)
American	73	CNN (LV, AA), NBC (23, NW), MSNBC (11, 13)
Arabic	33	LBC (NW, NA, 2)
Chinese	52	CCTV (3, DY), NTDTV (12, 19)

Table 6.4: The news broadcasters and the total length of news videos in each ideology in the TRECVID'05 video archive. The different news channels from the same broadcasters are listed in the parentheses.

We identified 10 international news events in late 2004 and news videos covering these news events. The number of news stories on 10 news events from each ideological perspective is listed in Table 6.5. We automatically determined whether a news story was about a news event by checking whether a news event's keywords were spoken in the video's automatic speech recognition transcripts. For the non-English news programs, the TRECVID organizer, NIST, provided the English translations.

We used the visual concept annotations from the Large-Scale Concept Ontology for Multimedia (LSCOM) v1.0 (Kennedy & Hauptmann, 2006) (see Section 3.2.2).

We conducted the experiments first using the LSCOM annotations, and later replaced manual annotations with predictions from empirically trained visual concept classifiers. Using manual annotations is equivalent to assuming that we have very accurate visual

News Event	American	Arabic	Chinese
Peace summit	13	10	16
Cabinet changes	39	-	13
Mideast peace	17	-	-
Iraq War	65	66	100
Presidential election	41	24	49
Arafat's death	118	125	65
AIDS	-	-	11
Afghanistan situation	15	-	19
Powell's resignation	28	12	-
Iranian nuclear weapons	-	17	36

Table 6.5: The number of news stories about a news event reported from each ideological perspective. If the number of news stories about a news topic from a particular perspective is fewer than 10, they are marked as "-".

concept classifiers. Given that state-of-the-art classifiers for most visual concepts are far from perfect, why would we want to start from manual annotations assumed to be perfect concept classifiers?

- First, manual visual concept annotations enable us to test our idea without being confounded by the poor accuracy of visual concept classifiers. If we started from poor concept classifiers and found that our idea did not work, we could not know whether a) our idea indeed cannot identify a news video's ideological perspective or b) the idea could work but the classifiers' accuracy was too low.

- Second, manual annotations established the performance upper-bound of our method. We could relax the assumption by gradually injecting noise into manual annotations to decrease classifiers' accuracy until the accuracy reached the state of the art (see Section 6.2.3.5). We could thus have both realistic and optimistic pictures of what our method can achieve.

6.2.3.2 Identifying News Video's Ideological Perspective

We evaluated the idea of using emphatic patterns of visual concepts to identify a news video's ideological perspective in a classification task. For each ideological perspective, we trained a one-against-all binary classifier (e.g., the American perspective vs. non-American perspectives) using Joint Topic and Perspective Model (see Section 4.4). We then evaluated the performance of the ideological perspective classifier on held-out news videos. We compared the perspective classifiers based on the Joint Topic and Perspective Model with a random baseline (i.e., predicting one of the two perspectives with equal probabilities).

We conducted a total of 22 binary perspective classification experiments, and calculated the average F1 for each ideological perspective. The positive data of a perspective classification experiment consist of videos on the same news topic from a particular ideological perspective (e.g., news stories about "Arafat's death" from Arabic news broadcasters). The negative data consist of news videos on the same news topic but from contrasting ideological perspectives (e.g., news stories about "Arafat's death" from non-Arabic news broadcasters, that is, American plus Chinese news broadcasters). We discarded the news topic and ideological perspective combination in Table 6.5 that contained fewer than 10 news stories. We conducted 10-folded cross-validation in each binary classification task. We also varied the amount of training data from 10% to 90%.

We adopted the commonly used evaluation metrics for binary classification tasks: precision, recall, and F1 (Manning et al., 2008). Precision is the fraction of the predicted positive news stories that are indeed positive. Recall is the fraction of all positive news stories that are predicted positive. F1 is the geometric average of precision and recall. The random baseline's F1 may not be 0.5 because the proportion of positive and negative data is not the same in our data.

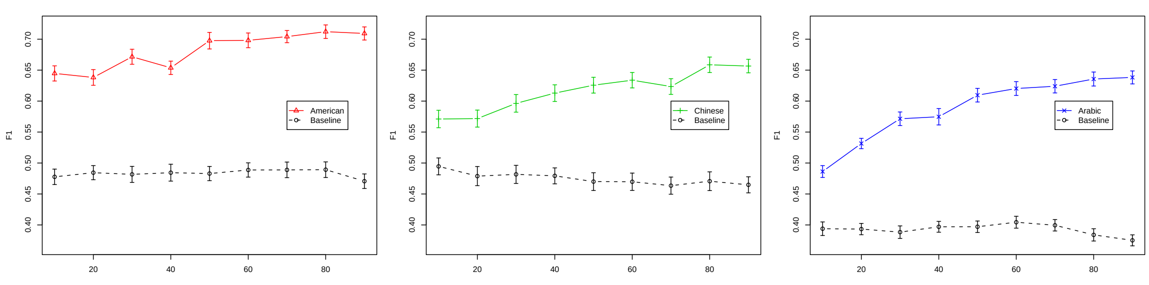


Figure 6.5: The experimental results of classifying a news video's ideological perspectives. The x axis is the amount of training data, and the y axis is the average F1.

We plot the classification results in Figure 6.5. The ideological perspective classifiers based on the Joint Topic and Perspective Model significantly outperformed the random baselines in three ideological perspectives. American perspective classifiers achieved the best performance of average F1 around 0.7. There were, however, more American news stories available for training than Arabic and Chinese perspectives.

The significantly better-than-random classification performance can be attributed to

- Emphatic patterns of visual concepts: News broadcasters holding different ideological beliefs seem to exhibit strongly and consistently emphatic patterns of visual concepts when they cover international news events. Therefore, by modeling the emphatic patterns of visual concepts, our classifiers can identify the ideological perspective from which a news video was produced.

- The Joint Topic and Perspective Model: The model seems to closely capture the emphatic patterns of visual concepts. The model assumptions (e.g., the multiplicative relation between topical and ideological weights, normal priors, etc.) do not seem to contradict real data much.

- Sufficient coverage of LSCOM: The visual concepts in the LSCOM ontology seem very extensive, at least in terms of covering news events in the TRECVID 2005 archive. Although LSCOM was initially developed to support video retrieval (M. Naphade et al., 2006), LSCOM seems to cover a wide variety of visual concepts so that the choices made by news broadcasters holding different ideological perspectives can be closely captured.

Since a news video's broadcaster is known when it was first recorded, is the task of identifying its ideological perspective as trivial as looking up its broadcaster from metadata?

- What is the most interesting in the experiments is not the perspective classification task *per se*. It is that our classifiers based *only* on emphatic patterns of visual concepts can significantly outperform random baselines.

- We can identify a news video's ideological perspective by looking up metadata, but this approach is not applicable to videos that contain little or no metadata (e.g., user-generated images on Flicker or videos on YouTube). We are more interested in a method of broad generalization, and thus have chosen to develop our method solely based on visual content and generic visual concepts without assuming the existence of rich metadata.

- So far, very few test beds exist for identifying a video's ideological perspective. The clearly labeled news videos in the TRECVID 2005 video achieve allow us to conduct controlled experiments.

6.2.3.3 Topical and Ideological Weights

We illustrated the emphatic patterns of visual concepts by visualizing the topical and ideological weights recovered from the news videos in the TRECVID 2005 archive. We explicitly modeled the emphatic patterns of visual concepts as a product between a concept's *topical* weight (i.e., how frequently a visual concept is shown for a specific news event) and *ideological* weights (i.e., how much emphasis a broadcaster holding a particular ideological perspective puts on it). These topical and ideological weights succinctly summarize the emphatic patterns of visual concepts.

We visualized the topical and ideological weights of visual concepts in a color text cloud. Text clouds, or tag clouds, have been a very popular way of displaying a set of short strings and their frequency information (e.g., bookmark tags on Del.icio.us[1] and photo tags on Flicker[2]). Text clouds represent a word's frequency in size, i.e., the value of topical weights τ in the model. The larger a word's size, the more frequently the word appears in a collection.

To show a visual concept's ideological weight, we painted a visual concept in color shades. We assigned each ideological perspective a color, and a concept's color was determined by which perspective uses a concept more frequently than the other. Color shades gradually change from pure colors (strong emphasis) to light gray (almost no emphasis). The degree of emphasis is measured by how far away a concept's ideological weight ϕ is from 1. Recall that when a concept's ideological weight ϕ is 1, it places no emphasis.

We fitted the Joint Topic and Perspective Model (Section 6.2.2) on the news videos about a specific news event (see Table 6.5) from two contrasting ideologies, (e.g., American vs. non-American, i.e., Chinese plus Arabic). For example, Figure 6.6 shows the topical weights (in word sizes) and ideological weights (in color shades) of the news stories about the Iraq War. The visual concepts of low topical and ideological weights are omitted due to space limit.

In reporting the Iraq War news, Figure 6.6 shows how American and non-American (i.e., Chinese and Arabic) news media presented stories differently. Concepts such as Outdoor, Adult, and Face were frequently shown (see Figure 6.2 and Figure 6.7), but they were not shown more or less frequently by different ideologies. Compared to non-American news broadcasters, American news media showed more battles (Fighter_Combat, Street_Battle), war zones (Exploding_Ordnance, Explosion_Fire, Smoke, Shooting), soldiers (Military_Personnel, Armed_Person, Police_Private_Security_Personnel), and weapons (Machine_Guns, Weapons, Rifles). In contrast, non-American news media showed more non-American people (Muslims, Asian_People) and symbols (Non-US_National_Flags),

[1] http://del.icio.us

[2] http://www.flickr.com/

Figure 6.6: The color text cloud summarizes the topical and ideological weights uncovered in the news videos about the Iraq War. The larger a word's size, the larger its topical weight. The darker a word's color shade, the more extreme its ideological weight. Red represents the American ideology, and blue represents the non-American ideologies (i.e., Arabic and Chinese).

and civilian activities (Parade, Riot). Although some visual concepts were emphasized in a manner that defied our intuition, the military vs. non-military contrast was clearly shown in how Western and Eastern media covered the Iraq War.

We show how Arabic news media and non-Arabic (i.e., Chinese and American) news media covered Arafat's death in Figure 6.7. We can see that Arabic news media reported more reactions from Palestinian people (People_Crying, Parade, Demonstration_-Or_Protest, People_Marching) as we suspected in Section 6.2.1. In contrast, non-Arabic news media showed more still images (Still Image) of Yasser Arafat (Yasser_Arafat) and reactions from political leaders (Head_Of_State, George_Bush). Again, we observe how news broadcasters holding contrasting ideological perspectives choose to emphasize different visual concepts.

An alternative way of estimating how frequently visual concepts are chosen is to obtain maximum likelihood estimates of a unigram language model (Manning & Schütze, 1999). There are also alternative ways of estimating what visual concepts are emphasized by each ideology (e.g., chi-square test, mutual information, etc. (Manning & Schütze, 1999)). The Joint Topic and Perspective Model differs from these techniques in the following aspects:

- Our model provides a probability model that unifies topical and ideological weights in the same model. Most of the previous techniques answer only one aspect of the question. The statistical model allows us to learn parameters and infer a news

Computers Protesters Sunny Corporate_Leader Attached_Body_Parts Sidewalks Host People_Crying Police_Private_-Security_Personnel Dresses_Of_Women Guard Smoke Parade Reporters Microphones Security_Checkpoint Windy Funeral Police Explosion_Fire Apartment_Complex Residential_Buildings Walking Office Road Guest Adobehouses Conference_Room Scene_Text Female_Reporter Muslims Car Demonstration_Or_Protest Exiting_Car Trees Vegetation Rocky_Ground Building Dirt_Gravel_Road Armed_Person Military_Personnel Flags Athlete Grandstands_Bleachers Sky Truck Urban_Scenes Computer_Or_Television_Screens People_Marching Streets Beards Exploding_Ordinance Nighttime Airport Congressman Celebration_Or_Party Suburban Highway Single_Family_Homes Handshaking Sports Politics Cityscape Clouds Landscape Text_Labeling_People Election_Campaign Waterways Text_On_Artificial_Background Greeting Overlaid_Text Maps Waterscape_Waterfront George_Bush Us_Flags Caucasians Motorcycle Head_Of_State Asian_People Non-uniformed_Fighters Non-us_National_Flags

Figure 6.7: The text cloud summarizes the topical and ideological weights uncovered from the news videos about the Arafat's death. The larger a word's size, the larger its topical weight. The darker a word's color shade, the more extreme its ideological weight. Red represents Arabic ideology, and blue represents non-Arabic ideologies (i.e., American and Chinese).

video's ideological perspective in a very principled manner.

- Our model explicitly models the emphatic patterns of visual concepts as a multiplicative relationship. The assumption may be arguably naive, but the concrete relationship allows future work for refinement. On the contrary, most of the previous techniques do not explicitly model how visual concepts are emphasized.

6.2.3.4 Does the Model Capture Ideological Perspectives or Production Styles?

We attributed the encouraging perspective classification results in Section 6.2.3.2 to the emphatic patterns of visual concepts, but the non-trivial classification performance can be also attributed to news broadcasters' production styles. Although we have removed non-news segments and news studio scenes, individual news broadcasters may still have idiosyncratic ways of editing and composing news footage. These news channel-specific product styles may be reflected in the visual concepts, and the high accuracy classification results in Section 6.2.3.2 may be mostly due to production styles and have little to do with ideological perspectives.

We tested the theory in the following classification experiment. Similar to the ideological perspective experiments in Section 6.2.3.2, we conducted classification experiments in a one-against-all setting (e.g., positive data are Arabic news stories, and negative data are the combined Chinese and American news stories) with a key difference: we did not contrast news stories on the *same* news event. If the Joint Topic and Perspective Model captured only individual news broadcasters' production styles, we would expect the classifiers to perform well in this new setting, no matter whether we contrasted news stories on the same news topic or not. Production styles should exist independent of news events.

We conducted three ideological classification experiments. For each ideology, we randomly sampled positive data from all possible news events in Table 6.5, and randomly sampled negative data from the news stories from the other two ideologies. For example, in classifying Chinese ideology, we collected positive data by randomly sampling Chinese news stories (about any news events), and negative data by randomly sampling from Arabic and American news stories (also without regard to their news topics). We trained the perspective classifiers based on the Joint Topic and Perspective Model, and performed 10-fold cross-validation. We also varied the amount of training data from 10% to 90%. We compared the perspective classifiers with random baselines (i.e., randomly guessing one of two perspectives with equivalent probabilities).

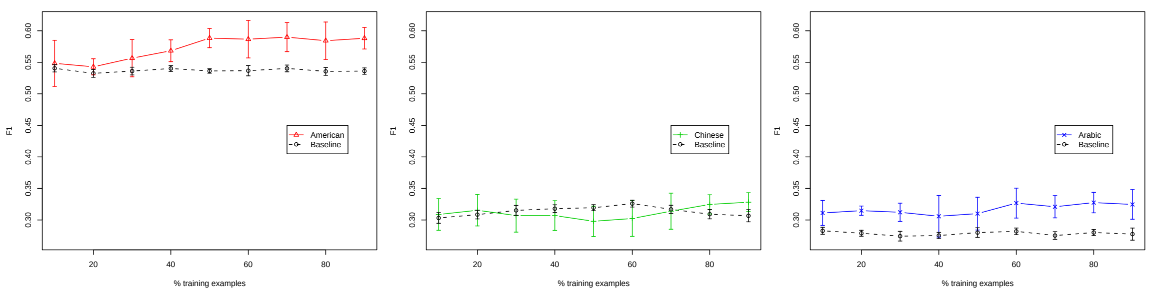


Figure 6.8: The experimental results of testing the theory that the Joint Topic and Perspective Model captures only individual news broadcasters' production styles but not emphatic patterns of visual concepts. The x axis is the amount of training data. The y axis is the average F1.

We plot the experimental results in Figure 6.8. Except for Chinese ideology, there are statistically significant differences when training data are large ($p < 0.01$). Therefore, the classifiers seemed to recognize some production styles, at least in American and Arabic news stories, that allow classifiers to outperform random baselines. However, the difference is minor, and much smaller than the difference we observed in Figure 6.5 when news stories were contrasted on the *same* news event. Therefore, individual broadcasters' production styles contribute to but cannot account for the high accuracy of the perspective classification results in Figure 6.5. In addition to a minor effect from production styles, broadcasters holding different ideological perspectives seemed to exhibit strongly emphatic patterns of visual concepts when they covered international news events. By exploiting these emphatic patterns we can successfully identify the ideological perspective from which a news video was made.

6.2.3.5 Visual Concept Classifiers' Accuracy

So far our experiments have been based on manual annotations of visual concepts from LSCOM. Using manual annotation is equivalent to assuming that perfect concept classifiers are available, which is not practical given that the state-of-the-art classifiers for most visual concepts are still far from perfect (M. R. Naphade & Smith, 2004). So, how well can we actually use the emphatic patterns of visual concepts to identify a news video's ideological perspective if we use empirically trained classifiers?

We empirically trained all LSCOM visual concept classifiers using Support Vector Machines. For each concept, we first trained uni-modal concept classifiers using many low-level features (e.g., color histograms in various grid sizes and color spaces, texture, text, audio, etc.), and then built multi-modal classifiers that fused the outputs from top uni-modal classifiers (see (A. G. Hauptmann et al., 2005) for more details about the training procedure). We obtained a visual concept classifier's empirical accuracy by training on 90% of the TRECVID 2005 development set and testing on the held-out 10%. We evaluated the performance of the best multi-modal classifiers on the held-out set in terms of average precision.

We varied visual concept classifiers' accuracy by injecting noise into manual annotations. We randomly flipped the positive and negative LSCOM annotations of a visual concept until we reached the desired break-even points of recall and precision. Simulating a set-based evaluation metric (e.g., recall-precision break-even point) is easier than simulating a rank-based evaluation metric (e.g., average precision). Recall-precision break-even points are shown to be highly correlated with average precision (Manning et al., 2008). We varied the classifiers' break-even points ranging from average precision obtained from empirically trained classifiers to 1.0 (i.e., the original LSCOM annotations), and repeated the perspective classification experiments in Section 6.2.3.2. The noise injection allowes us to easily manipulate a classifier's accuracy, but the real classification errors may not be completely random. The following results should be interpreted with a grain of salt.

The experimental results in Figure 6.9 show that using the empirically trained visual concept classifiers (the leftmost data points) still outperformed random baselines in identifying Arabic, Chinese, and American ideological perspectives (t-test, $p < 0.01$). The improvement, however, is smaller than that found by using manual LSCOM annotations (the rightmost data points). The median average precision of the empirically trained classifiers for all LSCOM concepts was 0.0113 (i.e., the x coordinate of the leftmost data

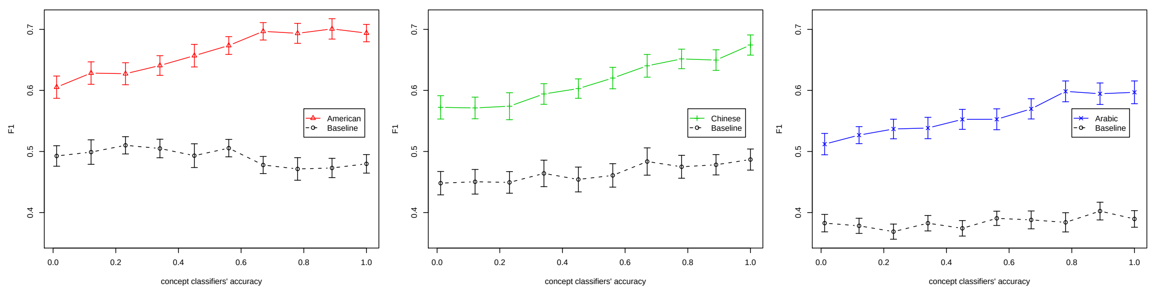


Figure 6.9: The experimental results of varying visual concept classifiers' accuracy. The x axis is the varied concept classifier's accuracy in terms of recall-precision break-even point. The leftmost data point is the experiment using empirically trained visual concept classifiers. The rightmost data point is the experiment using perfect visual concept classifiers, i.e., LSCOM manual annotations.

point in Figure 6.9). Not surprisingly, perspective identification improved as the concept classifiers' performance increased.

We should not be easily discouraged by current classifiers' poor performance. With the advances in computational power and statistical learning algorithms, it is likely that concept classifiers' accuracy will be continuously improved. Moreover, we may be able to compensate for concept classifiers' poor accuracy by increasing the number of visual concepts, as suggested in a recent study of significantly improving video retrieval performance using thousands of visual concepts (A. Hauptmann et al., 2007).

6.3 Identifying Ideological Perspectives in Web Videos

Although researchers have made great advances in automatically detecting "visual concepts" (e.g., car, outdoor, and people walking) (M. R. Naphade & Smith, 2004), developing classifiers that can automatically identify whether a video express a pro-life view on the abortion issue is still a very long-term research goal. The difficulties inherent in content-based approaches may explain why the problem of automatically identifying a video's ideological perspective on an issue has received little attention.

- In this section we study identifying a web video's ideological perspective on political and social issues using associated tags. We have previously shown that individual news broadcasters' bias can be reliably identified based on a large number of visual concepts (Lin & Hauptmann, 2008). We show that ideological perspectives are not only reflected in the selection of visual concepts but also in tags describing the content of videos.

Videos on video sharing sites such as YouTube allow users to attach tags to categorize and organize videos. The practice of collaboratively organizing content by tags is called folksonomy, or collaborative tagging. In Section 6.3.2.2 we show that a unique pattern of tags emerging from videos expressing opinions on political and social issues.

- In Chapter 4, we apply a statistical model to capture the pattern of tags from a collection of web videos and associated tags. The statistical model simultaneously captures two factors that account for the frequency of a tag associated with a web video: the subject matter of a web video and the ideological perspective that a video's author takes on an issue.
- We evaluate the idea of using associated tags to classify a web video's ideological perspective on an issue in Section 6.3.2. The experimental results in Section 6.3.2.1 are very encouraging, suggesting that Internet users holding similar ideological beliefs upload, share, and tag web videos similarly.

6.3.1 Joint Topic and Perspective Models for Web Videos

We apply a statistical model to capture how web videos expressing a particular ideological perspective are tagged. The statistical model, called the Joint Topic and Perspective Model (see Chapter 4), is designed to capture an *emphatic* pattern of words in ideological discourse. In the context of web videos, we apply the model to capture the emphatic patterns of *tags* associated with web videos.

The emphatic pattern consists of two factors: *topical* and *ideological*. For example, in the web videos on the abortion issue, tags such as abortion and pregnancy are expected to occur frequently no matter what ideological perspective a web video's author takes on the abortion issue. These tags are called *topical*, capturing what an issue is about. In contrast, the occurrences of tags such as pro-life and pro-choice vary depending on a video author's view on the abortion issue. These tags are emphasized (i.e., tagged more frequently) on one side and de-emphasized (i.e., tagged less frequently) on the other side. These tags are called *ideological*.

The Joint Topic and Perspective Model assigns *topical* and *ideological* weights to each tag. The topical weight of a tag captures how frequently the tag is chosen because of an issue. The *ideological* weight of a tag represents to what degree the tag is emphasized by a video author's ideology on an issue. The Joint Topic and Perspective Model assumes that the frequency of a tag is associated with how a web video is governed by these two sets of weights.

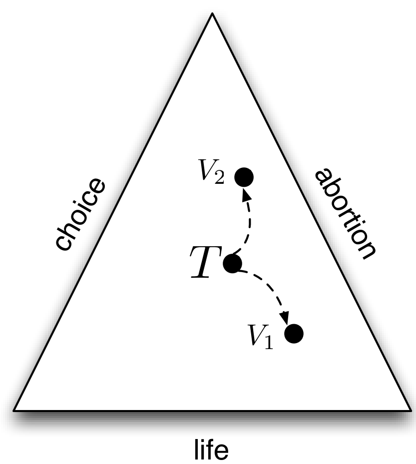


Figure 6.10: A three-tag simplex illustrates the main idea of the Joint Topic and Perspective Model for web videos. T denotes the proportion of the three tags (i.e., topical weights) that are chosen for a particular issue (e.g., abortion). V_1 denotes the proportion of the three tags after the topical weights are modulated by video authors holding the "pro-life" view; V_2 denotes the proportion of the three tags modulated by video authors holding the contrasting "pro-choice" view.

We illustrate the main idea of the Joint Topic and Perspective Model in the context of web videos using a three-tag world in Figure 6.10. Any point in the three-tag simplex represents the proportion of three tags (i.e., abortion, life, and choice) chosen in web videos about the abortion issue (also known as a multinomial distribution's parameter). T represents how likely we would be to see abortion, life, and choice in web videos about the abortion issue. Suppose a group of web video authors holding the "pro-life" perspective choose to produce and tag more life and fewer choice. The *ideological* weights associated with this "pro-life" group move the proportion from T to V_1. When we sample tags from a multinomial distribution of a parameter at V_1, we would see more life and fewer choice tags. In contrast, suppose a group of web video authors holding the "pro-choice" perspective choose to make and tag more choice and fewer life. The *ideological* weights associated with this "pro-choice" group move the proportion from T to V_2. When we sample tags from a multinomial distribution of a parameter at V_2, we would see more life and fewer choice tags. The topical weights determine the position of T in a simplex,

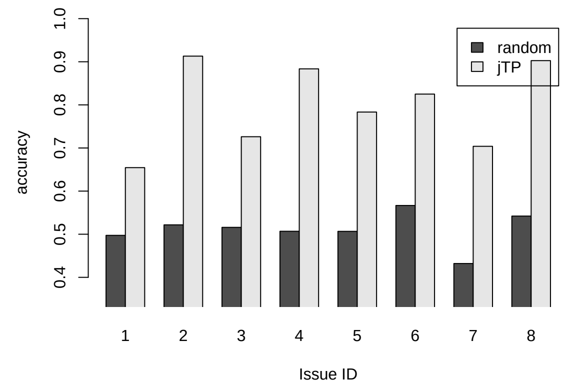


Figure 6.11: The accuracies of classifying a web video's ideological perspective on eight issues

and each ideological perspective moves T to a biased position according to its ideological weights.

We can fit the Joint Topic and Perspective Model on data to simultaneously uncover topical and ideological weights. These weights succinctly summarize the emphatic patterns of tags associated with web videos about an issue. Moreover, we can apply the learned weights from a training set, and predict the ideological perspective of a new web video based on associated tags.

6.3.2 Experiments

We collected web videos expressing opinions on various political and social issues from YouTube[3] (see Section 3.2.3).

6.3.2.1 Identifying Videos' Ideological Perspectives

We evaluated how well a web video's ideological perspective can be identified based on associated tags in a classification task. For each issue, we trained a binary classifier based on the Joint Topic and Perspective Model in Chapter 4, and applied the classifier on a held-out set. We calculated the average accuracy of the 10-fold cross-validation. We compared the classification accuracy using the Joint Topic and Perspective Model with a baseline that randomly guesses one of two ideological perspectives. The accuracy of a random baseline is close but not necessarily equal to 50% because the numbers of videos in each ideological perspective on an issue are not necessarily equal.

The experimental results in Figure 6.11 are very encouraging. The classifiers based on the Joint Topic and Perspective Model (labeled as jTP in Figure 6.11) outperform the random baselines for all eight political and social issues. The positive results suggest that the ideological perspectives of web videos can be identified using associated tags. Because the labels of our data are noisy, the results should be considered as a lower bound. The actual performance may be further improved if less noisy labels are available.

The positive classification results also suggest that Internet users sharing similar ideological beliefs on an issue appear to author, upload, and share similar videos, or at least, to tag similarly. Given that these web videos are uploaded and tagged at different times without coordination, it is surprising to see any pattern of tags emerging from folksonomies of web videos on political and social issues. Although the theory of ideology has argued that people sharing similar ideological beliefs use similar rhetorical devices for expressing their opinions in the mass media (Van Dijk, 1998), we are the first to observe this pattern of tags in user-generated videos.

The non-trivial classification accuracy achieved by the Joint Topic and Perspectives Model suggests that the statistical model closely matches with the real data. Although the Joint Topic and Perspective Model makes several modeling assumptions, including a strong assumption on the independence between tags (through a multinomial distribution), the high classification accuracy supports that these assumptions do not violate the real data too much.

6.3.2.2 Patterns of Tags Emerging from Folksonomies

We illustrate the patterns of tags uncovered by the Joint Topic and Perspective Model in Figure 6.12 and Figure 6.13. We show tags that occur more than 50 times in the collection. Recall that the Joint Topic and Perspective Model simultaneously learns the topical weights τ (how frequently a word is tagged in web videos on an issue) and ideological weights ϕ (how frequently a tag is emphasized by a particular ideological perspective). We summarize these weights and tags in a color text cloud, where a word's size is correlated with the tag's topical weight, and a word's color is correlated with the tag's ideological weight. Tags that are not particularly emphasized by either ideological perspective are painted light gray.

The tags with large topical weights appear to represent the subject matter of an issue. The tags with large topical weights on the abortion issue in Figure 6.12 include `abortion`, `pro life`, and `pro choice`, which are the main topic and two main ide-

[3] http://www.youtube.com/.

Figure 6.12: The color text cloud summarizes the topical and ideological weights learned in the web videos expressing contrasting ideological perspectives on the abortion issue. The larger a word's size, the larger its topical weight. The darker a word's color shade, the more extreme its ideological weight. Red represents the pro-life ideology, and blue represents the pro-choice ideology.

ologies. The tags with large topical weights on the global warming issue in Figure 6.13 include `global warming`, `Al Gore` and `climate change`. Interestingly, tags with large topical weights are not particularly emphasized by either of the ideological views on the issue.

The tags with large ideological weights appear to closely represent each ideological perspective. Users holding the pro-life beliefs on the abortion issue (red in Figure 6.12) upload and tag more videos about `unborn baby` and religion (`Catholic`, `Jesus`, `Christian`, `God`). In contrast, users holding the pro-choice beliefs on the abortion issue (blue in Figure 6.12) upload more videos about women's rights (`women`, `rights`, `freedom`) and atheism (`atheist`). Users who acknowledge the crisis of global warming (red in Figure 6.13) upload more videos about energy (`renewable energy`, `oil`, `alternative`), recycle (`recycle`, `sustainable`), and pollution (`pollution`, `coal`, `emissions`). In contrast, users skeptical about global warming upload more videos that criticize global warming (`hoax`, `scam`, `swindle`) and suspect it is a conspiracy (`NWO`, `New World Order`).

We do not intend to give a full analysis of why each ideology chooses and emphasizes these tags, but to stress that folksonomies of the ideological videos on the Internet are a rich resource to be tapped. Our experimental results in Section 6.3.2.1 and the analysis in this section show that by learning patterns of tags associated with web videos, we can identify web videos' ideological perspectives on various political and social issues with high accuracy.

Folksonomies mined from video sharing sites such as YouTube contain up-to-date information that other resources may lack. Due to the data collection time coinciding with the United States presidential election, many videos are related to presidential candidates

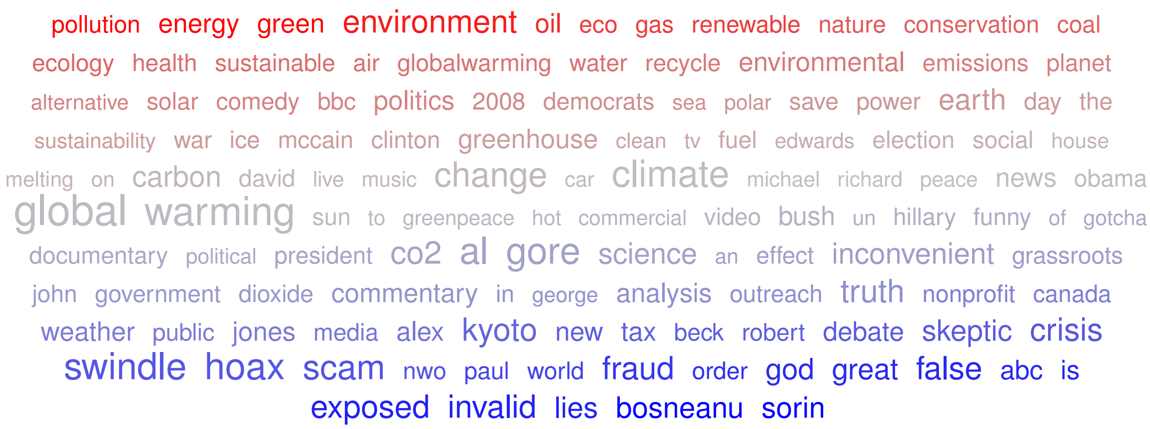


Figure 6.13: The color text cloud summarizes the topical and ideological weights learned in the web videos expressing contrasting ideological perspectives on the global warming issue. Red represents the ideology of global warming supporters, and blue represents the ideology of global warming skeptics.

and their views on various issues. The names of presidential candidates occur often in tags, and their views on various social and political issues become discriminative features (e.g., `Ron Paul`'s pro-life position on the abortion issue in Figure 6.12). Ideological perspective classifiers should build on folksonomies of web videos to take advantage of these discriminative features. Classifiers built on static resources may fail to recognize these current, but very discriminative, tags.

Chapter 7

Identifying Ideological Perspectives at the Sentence Level

In this chapter, we study the problem of identifying the sentences that clearly convey a particular ideological perspective. In previous chapters, we have shown that ideological perspectives can be reliably identified at the corpus and document level.

- We develop a hierarchical statistical model in Section 7.1 for ideological perspectives conveyed at both document and sentence levels. The model extends the Joint Topic and Perspective Model in Chapter 4 and adds a sentence layer. The model is called a Joint Topic and Sentence Perspective Model. To cope with the computationally intractable inference problem, we develop a variational inference algorithm in Section 7.1.2.

- To date, no human annotations on the opacity[1] of ideological perspectives are available. Annotating sentences clearly conveying an ideological perspective (i.e., less opaque) is difficult. We propose a method of annotating sentence-level opacity of

[1]Initially we used "intensity" instead of "opacity" to describe to what degree a sentence conveys an ideological perspective. A sentence clearly conveying an ideological perspective (of a document's author) was of high "intensity". However, the usage of "intensity" is ambiguous, as "intensity" is also used to indicate the strength that an ideological perspective, or subjectivity in general, is expressed. A sentence can be of low "intensity" in expression but of high "intensity" in conveying an ideological perspectives at the same time, for example, "I am a Democrat". To distinguish these two kinds of usages, we use the word "opacity" to describe to what degree a sentence conveys an ideological perspective, and save the word "intensity" to describe the strength of a subjective expression. We can thus more accurately characterize the above example with low "intensity" of expression and low "opacity." We thank Janyce Wiebe for kindly pointing out this distinction and suggesting to replace intensity here with opacity.

ideological perspectives by aggregating group judgments in Section 7.2. We evaluate the annotation method and assess its reliability in Section 7.2.2.

- We evaluate the Joint Topic and Sentence Perspective Model in Section 7.3. The experimental results are encouraging. The method can predict a sentence's opacity of conveying an ideological perspective better than two baselines.

7.1 A Joint Topic and Sentence Perspective Model

We develop a *hierarchical* model for ideological beliefs conveyed at the document and sentence levels. We have presented the Joint Topic and Perspective Model (jTP) that captures the ideological perspectives conveyed at the document level in Chapter 4. In this section, we extend jTP and add a sentence layer between documents and words to capture the ideological perspectives conveyed at the sentence level. We call this extended model a Joint Topic and Sentence Perspective Model (jTPs).

In jTPs, a document is associated with a random variable representing how likely the document is to convey a particular ideological perspective, and every sentence within the document is associated with a random variable representing how likely the sentence is to convey a particular ideological perspective. jTP ignores the sentence boundaries in text and assumes that every word conveys the same ideological perspective. In contrast, jTPs relaxes the assumption and allows some sentences to convey ideological perspectives in different degrees. jTPs can thus identify sentences that clearly convey a particular ideological perspective while jTP cannot.

We describe the jTPs in detail in Section 7.1.1, and develop a variational inference algorithm for jTPs in Section 7.1.2.

7.1.1 Model Specification

Formally, the Joint Topic and Sentence Perspective Model (jTPs) assumes the following generative process for an ideological text:

$$
\begin{aligned}
P_d &\sim \text{Bernoulli}(\pi), d = 1, \ldots, D \\
Q_{d,s} | P_d = v &\sim \text{Bernoulli}(\gamma_v), s = 1, \ldots, S_d, v = 1, \ldots, V \\
W_{d,s,n} | Q_{d,s} = v &\sim \text{Multinomial}(\beta_v), n = 1, \ldots, N_{d,s} \\
\beta_v^w &= \frac{\exp(\tau^w \times \phi_v^w)}{\sum_{w'} \exp(\tau^{w'} \times \phi_v^{w'})} \\
\tau &\sim \text{N}(\mu_\tau, \Sigma_\tau) \\
\phi_v &\sim \text{N}(\mu_\phi, \Sigma_\phi).
\end{aligned}
$$

The graphical model representation of the model can be seen in Figure 7.1.

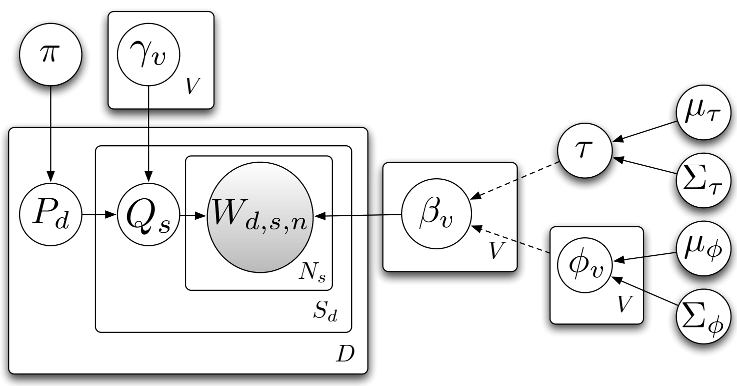


Figure 7.1: A Joint Topic and Sentence Perspective Model (jTPs) in a graphical model representation

In this thesis, we focus on bipolar political and social issues that are discussed mainly from two ideological perspectives, and thus the number of ideological perspective V is 2.

At the document level, jTPs represents the ideological perspective from which a document is written or spoken as a Bernoulli variable P_d, $d = 1, \ldots, D$, where D is the total number of documents in a collection. π is the parameter of the Bernoulli distribution for a document-level perspective P_d.

At the sentence level, jTPs represents the ideological perspective from which a sentence is written or spoken, again, as a Bernoulli variable $Q_{d,s}$, $s = 1, \ldots, S_d$, where S_d

is the total number of sentences in the d-th document in a collection. γ_v is the parameter of the Bernoulli distribution for a sentence-level perspective $Q_{d,s}$. Each word in the s-th sentence of the d-th document in a collection, $W_{d,s,n}$, is assumed to be sampled from a multinomial distribution conditioned on the sentence-level perspective $Q_{d,s}$ and document d's perspective, $n = 1, \ldots, N_s$, where N_s is the total number of words in a sentence.

In jTPs, we condition a sentence-level perspective $Q_{d,s}$ on its document-level perspective P_d, i.e., there is an edge between $Q_{d,s}$ and its parent P_d in the graphical representation of jTPs. We thus assume that the ideological perspective from which a sentence is written depends on the overall ideological perspective from which an article is written. We suspect that the relationship is one-sided, that is, γ_v is close to 0 or 1. For example, an author holding a pro-life stance may write more pro-life sentences than pro-choice sentences in an article. In contrast, an author holding a pro-choice stance may write more pro-choice sentences than pro-life sentences.

The parameter of a multinomial distribution, β_v^w, sub-scripted by an ideological perspective v and super-scripted by w-th word in the vocabulary, consists of two parts: topical weights τ, and ideological weights ϕ_v. β is an auxiliary variable, and is *deterministically* determined by (latent) topical τ, and ideological weights $\{\phi_v\}$. The deterministic relationship is shown as dashed lines. The prior distributions for topical and ideological weights are assumed to be normal distributions, which are conveniently chosen to model real values.

We combine topical and ideological weights through a logistic function. The multinomial distribution of words $W_{d,s,n}$ requires the parameter β_v to sum to one, and the logistic function assures us that the combined weights satisfy the sum-to-one requirement.

We assume that the relationship between topical and ideological weights to be multiplicative. Because one is the multiplicative identify, a word with an ideological weight of one (i.e., $\phi_v = 1$) means that it is not emphasized or de-emphasized.

The parameters in jTPS, denoted as Θ, include $\pi, \gamma_v, \mu_\tau, \Sigma_\tau, \mu_\phi, \Sigma_\phi$. Words $\{W_{d,s,n}\}$ and document-level perspectives $\{P_d\}$ are always observed, and sentence-level perspectives $\{Q_{d,s}\}$, topical weights τ, and ideological weights $\{\phi_v\}$ are not observed (i.e., latent variables).

The idea of in our model, multi-level and aggregate evidence, is also invented in other domains and tasks, for example, sentiment (McDonald, Hannan, Neylon, Wells, & Reynar, 2007). Their model they use undirected models, and have a different inference algorithms that use max-margin and Viterbi. However, the idea is similar: not every sentence convey the stuff, and by incorporating the context we can do better.

7.1.2 Variational Inference

The quantities of most interest in the Joint Topic and Sentence Perspective Model (jTPs) are (latent) sentence-level perspectives $\{Q_{d,s}\}$, topical weights τ, and ideological weights $\{\phi_v\}$. To infer these latent variables from a collection of documents about an issue, we calculate the following conditional probability:

$$
\begin{aligned}
&P(\{Q_{d,s}\}, \tau, \{\phi_v\} | \{W_{d,s,n}\}, \{P_d\}; \Theta) \\
\propto &P(\tau|\mu_\tau, \Sigma_\tau) \prod_{v=1}^{V} P(\phi_v|\mu_\phi, \Sigma_\phi) \prod_{d=1}^{D} P(P_d|\pi) \prod_{s=1}^{S_d} P(Q_{d,s}|\gamma_v) \prod_{n=1}^{N_s} P(W_{d,s,n}|Q_{d,s}, \tau, \{\phi_v\}) \\
= &\mathrm{N}(\tau|\mu_\tau, \Sigma_\tau) \prod_{v=1}^{V} \mathrm{N}(\tau|\mu_\phi, \Sigma_\phi) \prod_{d=1}^{D} \mathrm{Bernoulli}(P_d|\pi) \prod_{s=1}^{S_d} \mathrm{Bernoulli}(Q_{d,s}|\gamma_v) \\
&\prod_{n=1}^{N_s} \mathrm{Multinomial}(W_{d,s,n}|P_d, \beta),
\end{aligned} \tag{7.1}
$$

where $N(\cdot)$, $\mathrm{Bernoulli}(\cdot)$ and $\mathrm{Multinomial}(\cdot)$ are the probability density functions of multivariate normal, Bernoulli, and multinomial distributions, respectively.

The joint posterior probability distribution of $\{Q_{d,s}\}$, τ and $\{\phi_v\}$ in (7.1), however, is computationally intractable because of the non-conjugacy of the logistic-normal prior. We thus approximate the posterior probability distribution using a variational method (Jordan et al., 1999), and estimate the parameters of jTPs using variational expectation maximization (Attias, 2000).

By the Generalized Mean Field Theorem (Xing et al., 2003), we can approximate the joint posterior probability distribution of $\{Q_{d,s}\}$, τ and $\{\phi_v\}$ as the product of individual functions of $Q_{d,s}$, τ and ϕ_v:

$$
P(\{Q_{d,s}\}, \tau, \{\phi_v\} | \{P_d\}, \{W_{d,s,n}\}; \Theta) \approx q_\tau(\tau) \prod_{v} q_\phi(\phi_v) \prod_{d} \prod_{s} q_Q(Q_{d,s}), \tag{7.2}
$$

where $q_{Q_{d,s}}(Q_{d,s})$, $q_\tau(\tau)$, and $q_{\phi_v}(\phi_v)$ are the posterior probabilities of the sentence-level ideological perspectives, topical weights, and ideological weights conditioned on the random variables on their Markov blanket, respectively.

Specifically, q_ϕ is defined as follows:

$$
\begin{aligned}
q_\tau(\tau) =& P(\tau|\{W_{d,s,n}\}, \{Q_{d,s}\}, \{\langle\phi_v\rangle\}; \Theta) \\
\propto& P(\tau|\mu_\tau, \Sigma_\tau) \prod_v P(\{\langle\phi_v\rangle\}|\mu_\phi, \Sigma_\phi) P(\{W_{d,s,n}\}|\tau, \{\langle\phi_v\rangle\}, \{Q_{d,s}\}) \qquad (7.3) \\
\propto& \, \mathrm{N}(\tau|\mu_\tau, \Sigma_\tau) \prod_{n=1}^{N_s} \mathrm{Multinomial}(\{W_{d,s,n}\}|\{Q_{d,s}\}, \tau, \{\langle\phi_v\rangle\}), \qquad (7.4)
\end{aligned}
$$

where $\langle\phi_v\rangle$ is the GMF message based on $q_{\phi_v}(\cdot)$ (defined later in (7.6)). From (7.3) to (7.4) we drop the terms unrelated to τ.

Calculating the GMF message for τ from (7.4) is computationally intractable because of the non-conjugacy between multivariate normal and multinomial distributions. We follow a similar approach as in (Xing, 2005), and make a Laplace approximation of (7.4).

First, we represent the word likelihood $\{W_{d,s,n}\}$ in the following exponential form:

$$
P(\{W_{d,s,n}\}|\{Q_{d,s}\}, \tau, \{\phi_v\}) = \exp\left(\sum_v n_v(\tau \bullet \phi_v) - \sum_v n_v^T \mathbf{1} C(\tau \bullet \phi_v)\right),
$$

where $\bullet$ is element-wise vector product, n_v is a word count vector under a sentence-level ideological perspective v, $\mathbf{1}$ is a column vector of one, and C function is defined as follows:

$$
C(x) = \log\left(1 + \sum_{p=1}^{P} \exp x_p\right),
$$

where P is the dimensionality of the vector x.

Second, we expand C using Taylor series to the second order around $\hat{x}$ as follows:

$$
C(x) \approx C(\hat{x}) + \nabla(x)(x - \hat{x}) + \frac{1}{2}(x - \hat{x})^T H(\hat{x})(x - \hat{x}),
$$

where ∇ is the gradient of C, and H is the Hessian matrix of C.

Finally, we plug the second-order Taylor expansion of C back to (7.4), rearrange terms about τ, and obtain the approximation of $q_\tau(\cdot)$ as a multi-variate normal distribution

$\mathrm{N}(\tau|\mu^*, \Sigma^*)$ with a mean vector μ^* and a variance matrix Σ^* as follows:

$$\Sigma^* = \left(\Sigma_\tau^{-1} + \sum_v \langle\phi_v\rangle \downarrow H(\hat{\tau} \bullet \langle\phi_v\rangle) \rightarrow \langle\phi_v\rangle\right)^{-1}$$

$$\mu^* = \Sigma^* \left(\Sigma_\tau^{-1}\mu_\tau + \sum_v n_v \langle\phi_v\rangle - \sum_v n_v^T \mathbf{1} \nabla C(\hat{\tau} \bullet \langle\phi_v\rangle)\langle\phi_v\rangle \right.$$
$$\left. + \sum_v n_v^T \mathbf{1} \langle\phi_v\rangle \bullet (H(\hat{\tau} \bullet \langle\phi_v\rangle)(\hat{\tau} \bullet \langle\phi_v\rangle))\right),$$

where $\downarrow$ is a column-wise vector-matrix product, and $\rightarrow$ is a row-wise vector-matrix product. $\hat{\tau}$ is set as $\langle\tau\rangle^{(t-1)}$, where the superscript denotes the GMF message in the $t-1$ (i.e., previous) iteration. The Laplace approximation for the logistic-normal prior has been shown to be tight (Ahmed & Xing, 2007).

q_ϕ in (7.2) is defined as follows:

$$q_\phi(\phi_v) = P(\tau|\{W_{d,s,n}\}, \{Q_{d,s}\}, \{\langle\phi_v\rangle\}; \Theta)$$
$$\propto P(\tau|\mu_\tau, \Sigma_\tau) \prod_v P(\{\langle\phi_v\rangle\}|\mu_\phi, \Sigma_\phi) P(\{W_{d,s,n}\}|\tau, \{\langle\phi_v\rangle\}, \{Q_{d,s}\}) \tag{7.5}$$
$$\propto \mathrm{N}(\tau|\mu_\tau, \Sigma_\tau) \prod_{n=1}^{N_s} \mathrm{Multinomial}(\{W_{d,s,n}\}|\{Q_{d,s}\}, \tau, \{\langle\phi_v\rangle\}), \tag{7.6}$$

q_ϕ can be approximated á la q_τ as a multivariate normal distribution $N(\phi_v|\mu^\dagger, \Sigma^\dagger)$ with a mean vector $\mu^\dagger$ and a variance matrix $\Sigma^\dagger$ as follows:

$$\Sigma^\dagger = \left(\Sigma_\phi^{-1} + \sum_v \langle\tau\rangle \downarrow H(\langle\tau\rangle \bullet \hat{\phi}_v) \rightarrow \langle\tau\rangle\right)^{-1}$$

$$\mu^\dagger = \Sigma^\dagger \left(\Sigma_\phi^{-1}\mu_\phi + \sum_v n_v \langle\tau\rangle - \sum_v n_v^T \mathbf{1} \nabla C(\langle\tau\rangle \bullet \hat{\phi}_v)\langle\tau\rangle \right.$$
$$\left. + \sum_v n_v^T \mathbf{1} \langle\tau\rangle \bullet (H(\langle\tau\rangle \bullet \hat{\phi}_v)\langle\tau\rangle)\right),$$

where we set $\hat{\phi}_v$ as $\langle\phi_v\rangle^{(t-1)}$.

$q_Q(\cdot)$ in (7.2) is defined as follows,

$$q_Q(Q_{d,s}) = P(Q_{d,s}|\{W_{d,s,n}\}, P_d, \langle\tau\rangle, \{\langle\phi_v\rangle\}; \Theta) \tag{7.7}$$
$$\propto P(Q_{d,s}|P_d; \{\gamma_v\}) P(\{W_{d,s,n}\}|Q_{d,s}, \langle\tau\rangle, \{\langle\phi_v\rangle\}; \Theta)$$
$$= \text{Bernoulli}(Q_{d,s}|\gamma_v) \prod_n \text{Multinomial}(W_{d,s,n}|Q_{d,s}, \langle\tau\rangle, \{\langle\phi_v\rangle\}). \tag{7.8}$$

We can calculate the GMF message from q_Q (the expected values) by evaluating q_Q with two ideological views and re-normalize the values.

The E-step of the variational EM for jTPS is a message passing loop. We iterate over all q functions in (7.2), that is, (7.4), (7.6), and (7.8), until convergence. We monitor the change in the auxiliary variable β and stop when the absolute change is smaller than a pre-specified threshold.

In the M-step of the variational EM for jTPS, π can be maximized by taking the sample mean of $\{P_d\}$. γ_v can be similarly maximized by taking the sample mean of $\{Q_{d,s}\}$.

We monitor the expected word likelihood and stop the variational EM loop when the change of expected word likelihood is less than a pre-specified threshold.

7.2 Annotating Opacity of Ideological Perspectives

In this section, we describe a method of annotating the opacity of ideological perspectives conveyed at the sentence level. Annotations on the ideological perspectives at the document level are available (Lin et al., 2006); annotations at the sentence level, however, remain scarce. Not all sentences in a biased document convey the overall ideological perspective to the same degree, and manual annotations are needed.

Annotations on the sentence-level opacity of ideological perspectives will be very valuable for developing linguistic theories of ideological perspectives. The annotations will also be vital for evaluating computer programs that automatically extract clearly biased sentences.

Annotating opacity of ideological perspectives, however, is challenging. The common practice for annotation opacity is to quantize opacity into discrete categories. For example, we could potentially allocate three categories (Opaque, Medium, and Clear) for the Palestinian perspective and the Israeli perspective, plus one Neutral category, resulting in a total of seven categories. However, training annotators to agree on each of seven categories is not trivial.

Instead, we ask annotators to make a simple binary decision: is the sentence more likely to be written from the Israeli perspective or the Palestinian perspective? We then aggregate binary decisions over a large number of annotators. While individual annotators have different thresholds on opacity of ideological perspectives, a sentence conveying clearly a Palestinian perspective will be likely to be labeled as Palestinian perspective by most annotators. On the other hand, a sentence conveying a weak Palestinian perspective will have a mixture of the Israeli and Palestinian annotations. In other words, we use the agreement between annotators as a proxy to the underlying opacity of conveying an ideology perspective.

- A group of annotators labeled the opacity of ideological perspectives conveyed in 250 sentences extracted from the web articles on the Israeli-Palestinian conflict (Section 7.2.2.1).

- We quantitatively measured opacity of ideological perspectives by aggregating binary judgments from a group of annotators (Section 7.2.1). Intuitively, clearly one-sided sentences would be more likely to be consistently labeled by a majority of annotators, while a neutral sentence would be equally likely to be labeled as displaying either perspective. We call this annotation method the Vox Populi Annotation.

- We want to ensure the reliability of Vox Populi Annotation method. How many annotators do we need to reliably estimate the opacity of the ideological perspective conveyed in a sentence (Section 7.2.1.1)? Are these opacity measures consistent across different groups of annotators (Section 7.2.1.2)?

7.2.1 Vox Populi Annotation

We propose to quantitatively measure the degree to which a sentence conveys an ideological perspective by aggregating group judgments. We ask a group of annotators to make a forced binary choice on a sentence's ideological perspective, coded 0 for Perspective A and 1 for the contrasting perspective B. A sentence's opacity is estimated to be the average of group judgments, ranging between 0 and 1. The larger the average value, the less opaque a sentence conveys Perspective B (and the more opaque the sentence conveys Perspective A). We call this annotation method Vox Populi Annotation, and call the measure as Vox Populi Opacity.

The Vox Populi Annotation method is easy to implement. To annotate a sentence's opacity in conveying a particular ideological perspective, Vox Populi Annotation instructions can be as simple as "Which side do you think the sentence was written from?". Compared with most annotation studies, Vox Populi Annotation requires very little annotator training. However, are these opacity measures using the Vox Populi Annotation method *reliable*?

The Kappa statistic (Carletta, 1996; Artstein & Poesio, 2005b) cannot adequately assess the reliability of Vox Populi Annotations because annotators are not expected to agree on the same sentence at all. Contrary to most annotation studies, the Vox Populi Annotation method expects a large number of annotators to agree *collectively*, not on an individual basis. A sentence of opacity 0.75 is expected to have a quarter of n annotators disagree with the other three quarters. If the Vox Populi Annotation method is indeed reliable, we may still expect to see considerable disagreements when the same sentence is labeled by a new group of n annotators, but the proportion should be close to 0.75.

Similar to the chance-corrected kappa statistic, we assess the reliability of Vox Populi Annotations by considering how many observed annotations can be attributed to random guessing.

- The Vox Populi Annotation method estimates a sentence's ideological opacity by aggregating group judgments, but how many annotators are needed to make reliable measurement? In Section 7.2.1.1 we quantify the exact relationship between the number of annotators and the desired reliability.

- After a group of annotators label a set of sentences, how do we assess whether these Vox Populi Opacities are random guesses? In Section 7.2.1.2, we propose a method of assessing the reliability of the Vox Populi Annotation method on a collection of sentences.

7.2.1.1 Number of Annotators

How many annotators do we need to be confident that Vox Populi Opacity is not random guessing? We see this question as a statistical testing problem, where the null hypothesis is $\mu = 0.5$, and the alternative hypothesis is $\mu \neq 0.5$, where μ is the mean of the opacity of a ideological perspective conveyed in a sentence. The Vox Populi Annotation method requires annotators to make forced binary choices for each sentence, and each choice is like flipping a coin, i.e., a Bernoulli experiment.

We choose the exact Binomial test (Conover, 1971) to test the above hypothesis. The test procedure depends on two factors: the number of annotators and a sentence's ideological opacity (i.e., μ). There is a trade-off between two factors. If a sentence is extremely one-sided (i.e., μ is very close to 0 or 1), we do not need many annotators to reject the null hypothesis that a sentence is randomly annotated. However, more annotators are needed if a sentence conveys a mild ideological perspective (i.e., μ is close to 0.5). Our confidence on the statistical testing procedure can be expressed as the p-value $p(x)$ of the exact binomial test, defined as follows:

$$p(x) = \sum_{i=0}^{x} \binom{n}{i} 0.5^n + \sum_{i=n-y}^{n} \binom{n}{i} 0.5^n,$$

where x is the number of annotators labeling a sentence as a particular perspective, n is the total number of annotators, and y is the number of integers between $\lceil n/2 \rceil$ and n whose binomial density under the null hypothesis is less than the density at x[2].

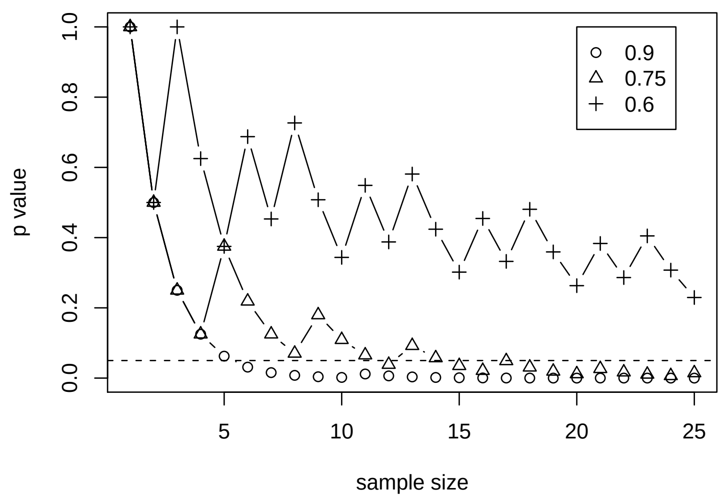


Figure 7.2: P-value decreases as an annotator group's size (sample size) increases. The horizontal dashed line is p-value 0.01. Three curves represent different Vox Populi Opacities. The curves zigzag due to the binomial distribution's discreteness.

We plotted the exact relationship between the number of annotators and and p-value for sentences of different Vox Populi Opacities in Figure 7.2. If we are to be confident when p-value is less than 0.01 (the dash line), a sentence of opacity 0.9 (or 0.1 due to the symmetry of the binomial distribution under the null hypothesis) requires six or more annotators (the x axis) to reject the hypothesis that the sentence is randomly annotated. A sentence of opacity 0.75 (or 0.25) needs more than 18 annotators to reject the null hypothesis. A sentence of opacity 0.6 requires more than 100 annotators (not shown in Figure 7.2). Generally, the more annotators, the more confident we are that Vox Populi Opacity is not random; the more intensely a sentence conveys an ideological perspective,

[2]We only list the case for $x < n/2$ and omit the case for $x \geq n/2$ because the two cases are very similar. See Conover (1971) for more details.

the fewer annotators we need to assess whether annotators make random guesses. By checking Figure 7.2 researchers can decide how many annotators are needed at the desired confidence level.

So far we have focused on testing whether one sentence is randomly annotated, but we usually have to apply a testing procedure repeatedly on multiple sentences. Suppose the desired confidence level is α. For each sentence, the probability of falsely treating a truly randomly annotated sentence as not randomly annotated (i.e., Type-I errors) is α. However, when we apply the same testing procedure on multiple sentences, the probability of making *at least one* Type-I error is much higher than α. This is the multiple testing problem.

Many methods have been proposed to alleviate the multiple testing problem (Shaffer, 1995; Dudoit, Shaffer, & Boldrick, 2003), and different ways of controlling Type-I errors have been proposed for the multiple testing problem, including Per-Comparison Error Rate, Family-Wise Error Rate, Per-Family Error Rate, and False Discovery Rate (Benjamini & Hochberg, 1995). By applying the above testing procedure described above to each sentence at a confidence level α, we can control Per-Comparison Error Rate at α.

Choosing Type-I error controls depends on our goal of annotating opacity of sentences and the percentage of truly randomly annotated sentences in an annotation study. Family-Wise Error Rate is the probability of at least one hypothesis test making Type-I error. To achieve low Family-Wise Error Rate, procedures such as the Bonferroni method declares very few sentences to be not randomly annotated. Family-Wise Error Rate is thus a very conservative measure and may work against the goal of discovering as many sentences that clearly conveying an ideological perspective as possible.

When large percentage of sentences are truly randomly annotated (because annotators are not well-trained or most sentences do not clearly convey an ideological perspective), the difference between Per-Comparison Error Rate and False Discovery Rate is very large. On the other hand, when large percentage of sentences conveying clearly an ideological perspective (i.e., not randomly annotated), the difference between Per-Comparison Error Rate and False Discovery Rate becomes minute (Dudoit et al., 2003).

7.2.1.2 Reliability

We assess the reliability of the Vox Populi Annotation method by assessing whether the Vox Populi Opacity from one group of annotators is similar to that from another group of annotators of the same size. The Vox Populi Annotation method is not reliable if opacity's magnitude changes greatly from one group of annotators to another group. Suppose 75% of annotators in one group label a sentence as Perspective A. The Vox Populi Annotation method is reliable if the same sentence is given to another group of annotators, and close to 75% of annotators still label the same sentence as Perspective A.

However, the above assessment method may be fooled by *random* guessing. Consider the following two random guessing cases. In the first case, annotators make completely random guesses between two contrasting perspectives, either because they are under-trained or ideological perspectives are too hard to identify at the sentence level. Either way, two groups of such random-guessing annotators will consistently output Vox Populi Opacity 0.5 for every sentence. The magnitude of opacities is similar, but it does not mean the annotations are not random.

In the second case, annotators keep making a biased decision, possibly because they are aware of the disproportion of two ideological perspectives in a corpus. Suppose two groups of annotators label a sentence to be the Israeli perspective 99% of the time. These two groups will label every sentence similarly, i.e., Vox Populi Opacity 0.99. However, we should not consider these Vox Populi Opacities as reliable because of superficial similarity resulting from biased guessing.

We choose Pearson's correlation coefficients to assess the reliability of the Vox Populi Annotation method. Given two sets of Vox Populi Opacities, $\{x_i\}$ and $\{y_i\}$, the Pearson correlation coefficient r is defined as follows:

$$r = \frac{n \sum_i x_i \sum_i y_i}{\sqrt{n \sum_i x_i^2 - (\sum_i x_i)^2} \sqrt{n \sum_i y_i^2 - (\sum_i y_i)^2}},$$

where n is the number of annotators. Correlation coefficients are positive and large when Vox Populi Opacity is positively correlated across different groups of annotators, and close to zero when Vox Populi Opacity is not related between different annotator groups. Correlation coefficients for the above two random cases will be zero because two groups of annotators make *independent* judgments (Casella & Berger, 2001). Therefore, the Vox Populi Annotation method is reliable if the correlation coefficients between two annotator groups are positive, high, and above zero.

We cannot use the κ statistic here (Carletta, 1996) because it considers mostly nominal and ordinal labels and does not handle quantitative labels well.

7.2.2 Measuring Opacity of Ideological Perspectives

7.2.2.1 Annotation corpus and procedure

We randomly chose 250 sentences from the bitterlemons corpus (Lin et al., 2006) (also see Section 3.1.1), which consists of articles published on the website http://bitterlemons.org/. The website is set up to "contribute to mutual understanding [between Palestinians and Israelis] through the open exchange of ideas[3]." Every week an issue about the

[3] http://www.bitterlemons.org/about/about.html

Israeli-Palestinian conflict is selected for discussion (e.g., “Disengagement: unilateral or coordinated?”), and a Palestinian editor and an Israeli editor each contribute one article addressing the issue. In addition, the Israeli and Palestinian editors invite one Israeli and one Palestinian to convey their views on the issue (sometimes in the form of an interview), resulting in a total of four articles in a weekly edition.

We recruited annotators from Carnegie Mellon University students and staff. Participants were asked to label a sentence in the recruitment advertisement. Each annotator signed a consent form that had been approved by the Institutional Review Board. A web-based interface displayed one sentence at a time, including the discussion topic and publication date. Annotators were instructed to judge the sentence by making a binary choice on the question “Do you think the sentence is written from the Israeli or Palestinian perspective?”. We encouraged participants to guess even when they were not sure. Eighteen of 26 participants finished the annotation study. Most of participants took one hour to finish the annotation study.

7.2.2.2 Annotation Results

Table 7.1 shows example sentences and their opacities of ideological perspectives.

The histogram of the Vox Populi Opacities of 250 sentences based on 18 annotators is shown in Figure 7.3. The opacity’s distribution is bimodal, with one peak around 0.35 (i.e., more Palestinian) and one around 0.65 (i.e., more Israeli). The bimodal distribution suggests that two ideological perspectives in the bitterlemons corpus seem to be identifiable at the sentence level. If ideological perspectives could not be identified at the sentence level, annotators would mostly make random guesses, resulting in a distribution of Vox Populi Opacities closely centered around 0.5.

The stretched distribution also suggests that our annotations contain sentences of varying opacities, which results in a language resource that is much richer than simply one-sided (i.e., all close to 0 or 1) or weak (i.e., all close to 0.5).

To cope with the multiple testing problem, we control two kinds of Type-I errors (see Section 7.2.1.1). When we chose to control Per-Comparison Error Rate at 0.05, there are 101 sentences (40.4%) that are not randomly annotated. When we choose to control False Discovery Rate at 0.05 using the Benjamini-Hochberg method (Benjamini & Hochberg, 1995), there are 40 sentences (16%) that are not randomly annotated. The adjusted p-value threshold is 0.0075, which is much lower than the nominal 0.05 and thus more sentences are considered as randomly annotated. Finally, when we choose to control Family-Wise Error Rate using the Bonferroni method, there are only 15 sentences (6%) that are not randomly annotated. The adjusted p-value threshold using the Bonferroni method is 0.0002, which is much lower than the nominal 0.05 and implies that most sentences are randomly

Opacity	Example
0.9231	The first is that Bush has placed the Palestinian-Israeli conflict squarely within the war against terrorism.
0.7693	This government, on the contrary, is trying in many ways, including through the so-called disengagement plan, to consolidate the occupation.
0.4616	This was not inevitable: as an election year approaches and the US sinks deeper into the Iraqi morass, Washington is simply not prepared to give high enough priority to the Israeli-Palestinian issue.
0.2308	Nor are security for Israelis and an end to terrorism – major topics of emphasis in Bush's presentation – likely to be achieved in this way.
0.0769	Palestinians in these ongoing debates have been basing their objections to the plan specifically on the argument that it contradicts the road map, for example on the issue of settlements.

Table 7.1: Five sentences and their Vox Populi Opacities of ideological perspectives. The larger the value, the more annotators judge a sentence to be written from the Israeli perspective.

annotated.

7.2.2.3 Reliability Assessment

We calculated the Pearson correlation coefficients[4] of Vox Populi Opacities between two annotator groups to assess reliability. As described in Section 7.2.1.2, given $2n$ annotators, we randomly divided them into two groups of n annotators. For example, if we have 12 annotators, we randomly divide them into two 6-people annotator groups. We then calculated the Vox Populi Opacities of 250 sentences from each group, and computed the correlation coefficients between these two sets of 250 Vox Populi Opacities. We repeatedly sample different $2n$ annotators, and calculated the average of the 100 correlation coefficients.

[4]The results using rank-based correlation methods (Kendall's tau and Spearman's rho) are similar and thus omitted.

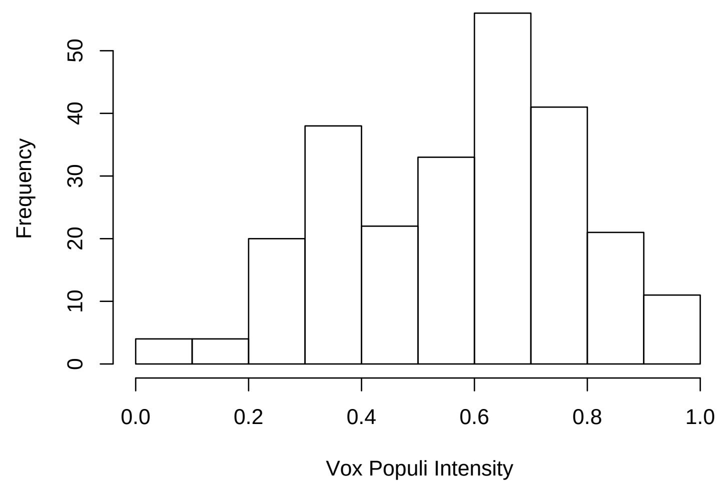


Figure 7.3: A histogram of Vox Populi Opacities of 250 sentences on the Israeli-Palestinian conflict. The larger the value, the more annotators judge a sentence to be written from the Israeli perspective.

We plot the correlation coefficients between two groups and two random guessing baselines in Figure 7.4. The correlation coefficients of the Vox Populi Annotation method differ significantly from 0 and two random baselines when the group size is large. The results suggest that Vox Populi Annotations, at least on the bitterlemons corpus, were unlikely to be randomly annotated and appear to be reliable. The positive correlation coefficients suggest that similar opacity estimates are likely to be obtained no matter where annotator groups come from. The median and maximal pair-wise kappa statistics among 18 annotators on the 250 sentences were 0.10 and 0.44, respectively, which are very unsatisfactory according to (Carletta, 1996). As group sizes get larger, the aggregated Vox Populi Opacity becomes positively and highly correlated between two annotator groups; even within each group, two annotators may still still disagree with each other (i.e., low pair-wise kappa).

7.2.3 Discussions

The Vox Populi Annotation method is not restricted to annotating to what degree a sentence conveys one of two contrasting perspectives. The Vox Populi Annotation method is

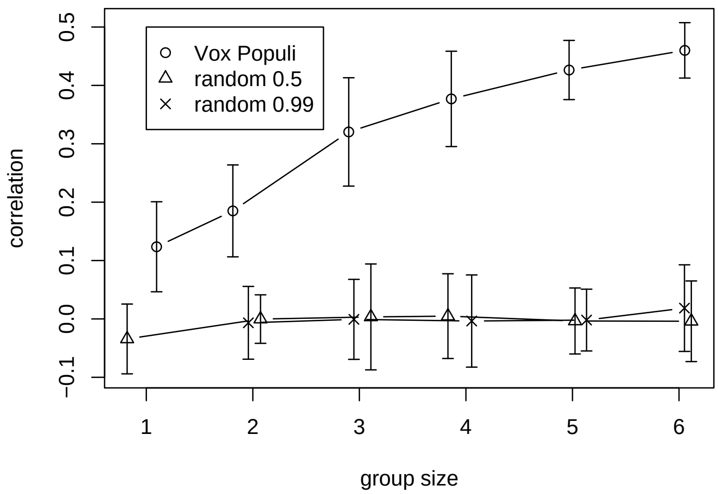


Figure 7.4: The correlation coefficients of Vox Populi Opacity and two random baselines as group sizes vary from one to six. We jittered the coordinate of group size to avoid the overlap between two random baselines.

a general methodology and may be applicable to other annotation tasks. As more computational linguists are interested in more complex linguistic phenomena (e.g., intensity of subjectivity (J. Wiebe et al., 2005), political and social controversies (Meyers, Ide, Denoyer, & Shinyama, 2007)), the Vox Populi Annotation method can be a viable alternative for researchers to quantitatively measure these complex phenomena.

However, the Vox Populi Annotation method is not applicable to annotation tasks that require extensive linguistic knowledge or have little ambiguities:

- Annotation tasks that require extensive linguistic knowledge include predicate argument structure in the Penn treebank (Meyers et al., 2007) and Rhetorical Structure Theory (RST) (Carlson, Marcu, & Ellen Okurowski, 2001). Because these annotation tasks require intensive training and constant monitoring, the cost of recruiting a large number of annotators becomes prohibitive. Besides, qualified candidates are very unlikely to be recruited from general public.

- Annotation tasks that have little ambiguities include named entities (Chinchor, 1997) and automatic speech recognition transcriptions (Fisher, Doddington, & Goudie-Marshall, 1986). Multiple annotators make little sense because they will all label very similarly.

Our annotation study is about labeling opacity of bipolar ideological perspectives, but how about some annotation tasks that require more than two choices? For example, an annotation study may investigate the ideologies of different ethnic groups on immigration issues, and ask a group of annotators to decide if a sentence is written from Asian, Hispanic, or African ethnic group's viewpoints (i.e., three categories). We can extend our reliability assessment in Section 7.2.1 to more than two choices. The exact binomial test for determining the number of annotators in Section 7.2.1.1 will be replaced by a multinomial test (Read & Cressie, 1988). The null hypothesis will not be a simple $\mu = 0.5$, and will be a multinomial vector that assumes that every category is equally likely. The correlation coefficient for assessing the reliability of the Vox Populi Annotation method will be replaced by multivariate correlation (DuBois, 1957).

There has been annotation studies on measuring intensity, for example, the intensity of opinioned expressions (J. Wiebe et al., 2005). The annotation schemes in previous work mostly use the Likert Scale (Likert, 1932) and quantize intensity into discrete categories (e.g., low, medium, strong, and extreme). To have two annotators agree on each scale requires extensive training. Moreover, it is not trivial to transform annotations in Likert Scales to numerical values (C.-H. Wu, 2007). On the contrary, Vox Populi Opacity is already a number and requires no transformation, which is important when evaluating computer programs that can output confidence scores.

The mathematical relationship between annotation group sizes and a sentence's opacity in Section 7.2.1.1 seems to be empirically observed. In a subjectivity annotation study (J. Wiebe et al., 2005),

> ... the difference between no subjectivity and a low-intensity private state might be highly debatable, but the difference between no subjectivity and a medium or high-intensity private sate is often much clearer.

The p-value formula based on the exact binomial test matches well the empirical observation. High opacity sentences are easier (i.e., requires fewer annotators) to be distinguished from random guessing than medium opacity sentences.

A large number of annotators had previously been used to reduce annotators' bias (Eugenio & Glass, 2004), that is, to minimize individual annotators' preferences to label on category more than the other category. Incidentally, 18 annotators, the same number as ours, were recruited in the study (Artstein & Poesio, 2005a). We explicitly determined the number of annotators based on the analysis in Section 7.2.1.1, and did not simply choose a big number.

One seeming obstacle to the Vox Populi Annotation method is the need for a large number of annotators. How can we afford so many annotators? While most annotation studies in computational linguistics recruit few annotators, many "annotation" tasks in other fields have begun to "recruit" a huge number of people, i.e., Crowd-sourcing (Hoew, 2006). Millions of Internet users have labeled web pages (e.g., Delicious[5]), photos (e.g., Flicker[6]), and videos (e.g., YouTube[7]) without being paid. ESP game (Ahn & Dabbish, 2004) and Google Image Labeler[8] use games to quickly collect high quality image annotations. With right kinds of incentive mechanisms and annotation platforms, annotation studies in computational linguistics are likely to replicate these success stories in other fields. The Vox Populi Annotation method is not for every annotation task, but for those annotation tasks that require little training, this thesis offers guidelines on selecting the number of annotators and assessing their reliability. Recently there has been an annotation study on sentiment conducted on Amazon Mechanical Turk (Barr & Cabrera, 2006), a commercial web service that facilitates large number of annotators.

7.3 Experiments

We evaluate the Joint Topic and Sentence Perspective Model (jTPs) on manually annotated sentences from editorials about the Israeli-Palestinian conflict. We compare the opacity of sentences estimated by jTPs and human annotations.

We divided the bitterlemons corpus into two sets: the testing set consists of 250 manually annotated sentences as described in Section 7.2, and the training set consists of all the sentences in the bitterlemons corpus (see Section 3.1.1) minus the 250 sentences in the testing set. We learned the parameters of jTPs on the training set, and predicted the opacities of the sentence in the test set.

To compare the predictions on sentence opacities and human annotations, we calculate absolute error (i.e., the absolute difference between two numerical values). Absolute error ranges between 0 and 1. The lower the absolute error, the more effectively jTPs is in uncovering the sentences that convey clearly a perspective.

We compare jTPs with two baselines. The first baseline (random1) is random guessing of two ideological perspectives (i.e., Israeli or Palestinian). The second baseline (random2) makes use of the information from document-level perspectives and predicts ideological perspective of a sentences to be the same as the perspective of the document containing the sentence (i.e., $Q_{d,s} = P_d$). For example, if a document is written from the Israeli perspective, the second baseline will always predict the ideological perspective of sentences from the document to be the Israeli perspective.

[5] http://del.icio.us/
[6] http://www.flickr.com/
[7] http://www.youtube.com/
[8] http://images.google.com/imagelabeler/

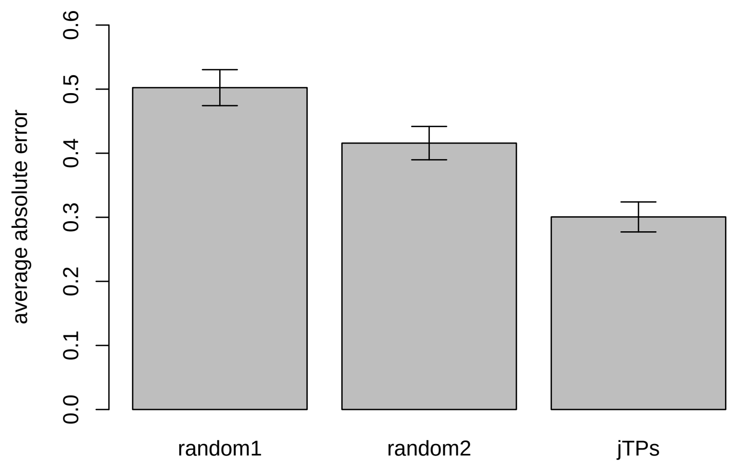


Figure 7.5: The average absolute error of sentence ideological opacity between manual annotations of 250 sentences and predictions from two baselines and the Joint Topic and Sentence Perspective Model (jTPs)

The experimental results in Figure 7.5 show that jTPs can successfully predict the opacity of a sentence conveying an ideological perspective. The predictions from jTPs are significantly closer (i.e., smaller absolute error) to manual annotations than two random baselines. The first baseline achieves close to 0.5 absolute error. By knowing the ideological perspective of the document, the second baseline achieves better performance than the first baseline. The result suggests that the ideological perspective of a sentence is likely to agree with the ideological perspective of the document containing the sentence.

Chapter 8

Conclusions

We investigated how ideological perspectives are reflected in text and video. When people discuss controversial issues, their choices of written words, spoken words, and visual concepts are not only about the topic (i.e., topical), but also reflect inner attitudes and opinions toward an issue (i.e., ideological). Although automatically understanding perspectives from written or spoken text documents and video is a scientifically challenging problem, it will enable many applications that can survey public opinion on social issues and political viewpoints on a much larger scale.

Machine understanding of subjective beliefs, however, has been deemed "a vain hope." (Abelson & Carroll, 1965) In this thesis, we took up the challenge and approached the problem in a statistical learning framework. The experimental results show that perspectives could be successfully identified by statistically modeling documents and their relationship.

- We presented a statistical model for ideological discourse. We hypothesized that ideological perspectives were partially reflected in an author or speaker's lexical choices. The experimental results show that the proposed Joint Topic and Perspective Model fit the ideological texts better than a model that naively assumes no lexical variations due to an author or speaker's ideological perspectives. We showed that the Joint Topic and Perspective Model uncovered words that represent an ideological text's topic as well as words that reveal ideological discourse structures. Lexical variations appeared to be a crucial feature that can enable automatic understanding of ideological perspectives from a large amount of documents.

- We developed a computational test of discerning different perspectives based on statistical distribution divergence between the statistical models of document collections. We showed that the proposed test can successfully separate document collections of different perspectives from other types of document collection pairs. The distribution divergence falling in the middle range can not simply be attributed to personal writing or speaking styles. From the plot of multinomial parameter difference, we offered insights into where the different patterns of distribution divergence come from.

- We presented a method of identifying different ideological perspectives expressed in broadcast news videos. We hypothesized that a broadcaster's ideological perspective was reflected in the composition of news footage. We showed that the visual concept based approach is effective in identifying news video pairs conveying different ideological perspectives as well as news video pairs about the same news event.

- We studied the problem of learning to identify the perspective from which a text is written at the document and sentence levels. We showed that much of a document's perspective is expressed in words and successfully uncovered the word patterns that reflect the author's perspective with high accuracy.

- We studied the problem of automatically identifying the ideological perspective from which a news video was produced. We presented a method based on specific, computable emphatic patterns of visual concepts: given a news event, contrasting ideologies emphasize different subsets of visual concepts. We explicitly modeled the emphatic patterns as a multiplicative relationship between a visual concept's topical and ideological weights, and developed an approximate inference algorithm to cope with the non-conjugacy of the logistic-normal priors. The experimental results suggest that the ideological perspective classifiers based on emphatic patterns are effective, and the high classification accuracy cannot be simply attributed to individual news broadcasters' production styles. Our work enables studying how video producers with different ideological beliefs convey their ideas and attitude in videos.

- We studied identifying the ideological perspective of a web video on an issue using associated tags. We showed that the statistical patterns of tags emerging from folksonomies can be successfully learned by a Joint Topic and Perspective Model, and the ideological perspectives of web videos on various political and social issues can be automatically identified with high accuracy. Web search engines and many Web 2.0 applications can incorporate our method to organize and retrieve web videos based on their ideological perspectives on an issue.

8.1 Future Directions

We identify several future directions for automatic ideology analysis. We made implicit modeling assumptions when we developed the Joint Topic and Perspective Model in Chapter 4, and we plan to relax these assumptions. We have shown that the emphatic patterns of words or visual concepts are informative of ideological perspectives. In the future, we plan to explore other patterns beyond lexical choices and the composition of visual concepts.

- Unsupervised Learning: We implicitly assume that documents and their ideological perspectives on an issue are available for training the Joint Topic and Perspective Model. However, the documents' ideological perspectives may not be available. Sometimes the documents' ideological perspectives may be very noisy.

  We would like to extend the Joint Topic and Perspective Model to a unsupervised setting, i.e., without assuming the existence of ideological perspective labels. Given a large text or video collection without any ideological perspectives, automatically determining if contrasting perspectives exist is a much more challenging problem. It is similar to a new way of clustering documents. Instead of clustering documents by topics as most clustering methods do (Steinbach, Karypis, & Kumar, 2000), an unsupervised ideology analysis will enable clustering by "ideological perspectives."

  One possible way to attacking the unsupervised ideological perspective clustering problem is to model the structure of the priors of ideological weights, i.e., μ_ϕ and Σ_ϕ. We can add a hyper prior to the prior of ideological weights, and learn the hyper prior distribution from multiple ideological collections. In other words, we can add another layer in the Joint Topic and Perspective Model. Instead of training on a corpus containing two contrasting ideological perspectives, we will train the extended model on multiple corpora, each of which containing two contrasting ideological perspectives.

  The unsupervised version of the Joint Topic and Perspective Model can be used to automatically identify if contrasting ideological perspectives exist in a corpora. News aggregation services can easily collect thousands of news stories on a news event every day, but it is very difficult to obtain ideological perspective labels for these news stories. The unsupervised Joint Topic and Perspective Model will be applicable in this case.

- Dynamics: We made an implicit assumption that ideological perspectives are known and static in the thesis. However, this may not be true for ideological perspectives on some issues. A news event may stay neutral but suddenly become controversial. Contrasting ideological perspectives emerge from stack-holders competing limit symbolic or physical resources. On the other hand, a controversy may end and two ideological perspectives may merge and become a single point of view. The dynamics of ideological perspectives should be taken into account.

We can extend the Joint Topic and Perspective model to capture the dynamics of lexical variations. Specifically, we can model the dynamics of the prior distributions of topical and ideological weights. Recent work on topic modeling has shown that dynamics of topics can be successfully uncovered (Ahmed & Xing, 2007; Blei & Lafferty, 2006; Nallapati, Ahmed, Xing, & Cohen, 2008).

- Beyond lexical choices: Van Dijk (1998) has identified numerous discourse structures that are commonly used in expressing ideological perspectives. We mainly exploit lexical variations in this thesis, but there are many discourse structures that remain to be incorporated from local syntax to global schematic structures (see Section 1.1). We expect the performance of identifying ideological perspectives to improve by incorporating these additional and complementary discourse structures. Furthermore, simultaneously considering multiple discourse structures will allow us to study the delicate interaction between these discourse structures used to express ideological perspectives.

  Metaphor has been shown to be indicative of the ideological perspective of an author on an issue. Lakoff (1992, 2003) have shown the "State-as-Person" metaphor that the pro-war ideology repeatedly used to justify wars. Lakoff (2003) illustrated the metaphor in the context of the Iraq War:

  > The war, we are told, is not being waged against the Iraqi people, but only against this one person. ... What this metaphor hides, of course, is that the 3000 bombs to be dropped in the first two days will not be dropped on that one person. They will kill many thousands of the people hidden by the metaphor, ...

  Many researchers have proposed methods of automatically identifying metaphors based on linguistic theories (Fass, 1991) and statistics from a large corpus (Mason, 2004) (also see a recent review paper (Zhou, Yang, & Huang, 2007)). So far these computational methods of identifying metaphors have been tested on simple, short sentences, and it is not clear how they will perform on ideological text. We plan to build on previous work and identify metaphors in ideological text. We will study the patterns of the metaphors used by authors holding different ideological perspectives. If authors from different ideological perspectives express their beliefs and value judgments through different use of metaphors, we can incorporate these patterns of metaphors that complement emphatic patterns of lexical choices in this thesis, and may further improve the performance of identifying ideological perspectives.

The usage of pronouns can be indicative of writers or speakers' ideological perspectives. Wilson (1990) analyzes the intricate use of pronouns in political speeches, and shows that the choices of pronouns reveal how much of the hidden implications in political discourse. In ideological discourse, the distinction between Us (people holding similar beliefs) and Them (people holding differing beliefs) are clearly made. How Us and Them are referred to in pronouns can reveal who writers or speakers determine group membership and thus their ideological perspectives on issues.

Appendix A

Gibbs Samplers for Modeling Individual Perspectives

Based the model specification described in Section 6.1.1 we derive the Gibbs samplers (Chen, Shao, & Ibrahim, 2000) for Latent Sentence Perspective Models as follows,

$$\pi^{(t+1)} \sim \mathrm{Beta}(\alpha_\pi + \sum_{n=1}^{N} d_n + \tilde{d}^{(t+1)},$$
$$\beta_\pi + N - \sum_{n=1}^{N} d_n + 1 - \tilde{d}^{(t+1)})$$

$$\tau^{(t+1)} \sim \mathrm{Beta}(\alpha_\tau + \sum_{n=1}^{N}\sum_{m=1}^{M_n} s_{m,n} + \sum_{m=1}^{\tilde{M}} \tilde{s_m},$$
$$\beta_\tau + \sum_{n=1}^{N} M_n - \sum_{n=1}^{N}\sum_{m=1}^{M_n} s_{m,n} + \tilde{M} - \sum_{m=1}^{\tilde{M}} \tilde{s_m})$$

$$\theta^{(t+1)} \sim \mathrm{Dirichlet}(\alpha_\theta + \sum_{n=1}^{N}\sum_{m=1}^{M_n} w_{m,n})$$

$$\Pr(S_{n,m}^{(t+1)} = s^1) \propto P(W_{m,n} | S_{m,n} = 1, \theta^{(t)})$$
$$\Pr(S_{m,n}^{(t+1)} = 1 | \tau, D_n)$$

$$\Pr(\tilde{D}^{(t+1)} = d^1) \propto \prod_{m=1}^{\tilde{M}} dbinom(\tau_d^{(t+1)})$$
$$\prod_{m=1}^{\tilde{M}} dmultinom(\theta_{d,\tilde{m}^{(t)}}) dbinom(\pi^{(t)})$$

where *dbinom* and *dmultinom* are the density functions of binomial and multinomial distributions, respectively. The superscript t indicates that a sample is from the t-th iteration. We run three chains and collect 5000 samples. The first half of burn-in samples are discarded. The Gibbs samplers are implemented in R (R Development Core Team, 2005).

References

Abelson, R. P. (1973). Computer models of thought and language. In (pp. 287–339). W. H. Freeman and Company. 2.1

Abelson, R. P., & Carroll, J. D. (1965, May). Computer simulation of individual belief systems. *The American Behavioral Scientist*, *8*, 24–30. 2.1, 2.1, 8

Ahmed, A., & Xing, E. P. (2007). On tight approximate inference of the logistic-normal topic admixture model. In *Proceedings of the eleventh international conference on artificial intelligence and statistics.* 4.2, 7.1.2, 8.1

Ahn, L. von, & Dabbish, L. (2004). Labeling images with a computer game. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 319–326). 7.2.3

Allan, J. (Ed.). (2002). *Topic detection and tracking: Event-based information organization*. Kluwer Academic Publishers. 2.6, 5.2.3.1

Amazon. (n.d.). *Amazon concordance.* `http://www.amazon.com/gp/search-inside/concordance-help.html`. 4.5.3

Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *Wired Magazine*, *16*(7). 2.1

Arango, M. (2004). Vanishing point. In *Proceedings of the twelfth ACM international conference on multimedia* (pp. 1067–1068). 2.6

Artstein, R., & Poesio, M. (2005a). Bias decreases in proportion to the number of annotators. In *Proceedings of the workshop on formal grammar – the mathematics of language* (pp. 141–150). 7.2.3

Artstein, R., & Poesio, M. (2005b, September). *Kappa*^{3} = *Alpha (or Beta)* (Tech. Rep. No. CSM-437). Department of Computer Science, University of Essex. 7.2.1

Attias, H. (2000). A variational bayesian framework for graphical models. In *Advances in neural information processing systems 12.* 4.2, 7.1.2

Barr, J., & Cabrera, L. F. (2006, May). AI gets a brain. *Queue*, *4*(4), 24–29. 7.2.3

Beineke, P., Hastie, T., & Vaithyanathan, S. (2004). The sentimental factor: Improving review classification via human-provided information. In *Proceedings of the 42nd conference on association for computational linguistics (acl).* 2.3

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. 7.2.1.1, 7.2.2.2

Berger, A. A. (1998). *Seeing is believing* (Second ed.). Mayfield Publishing Company. 1.2.5

Blei, D. M., & Lafferty, J. D. (2006). Correlated topic models. In *Advances in neural information processing systems (NIPS)*. 8.1

Blei, D. M., Ng, A. Y., & Jordan, M. (2003, January). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022. 2.5

Bocconi, S., & Nack, F. (2004). VOX POPULI: Automatic generation of biased video sequences. In *Proceedings of the first ACM workshop on story representation, mechanism and context* (pp. 9–16). 2.6

Buchanan, B. G., Barstow, D., Bechtal, R., Bennett, J., Clancey, W., Kulikowski, C., et al. (1983). Building expert systems. In (pp. 127–167). Addison-Wesley. 2.1

Carbonell, J. G. (1978). POLITICS: Automated ideological reasoning. *Cognitive Science*, *2*(1), 27–51. 2.1, 2.1, 2.6

Carbonell, J. G. (1979). *Subjective understanding: Computer models of belief systems*. Unpublished doctoral dissertation, Yale University. 2.1

Carletta, J. (1996). Assessing agreement on classification tasks: The kappa statistics. *Computational Linguistics*, *22*(2). 7.2.1, 7.2.1.2, 7.2.2.3

Carlson, L., Marcu, D., & Ellen Okurowski mary. (2001). Building a discourse-tagged corpus in the framework of rhetorical structure theory. In *Proceedings of the second SIGDial workshop on discourse and dialogue* (pp. 1–10). 7.2.3

Carruthers, S. L. (2000). *The media at war: Communication and conflict in the twentieth century*. St. Martin's Press. 1.2.1

Casella, G., & Berger, R. L. (2001). *Statistical inference* (Second ed.). Duxbury Press. 7.2.1.2

Chang, C.-C., & Lin, C.-J. (2001). *LIBSVM: a library for support vector machines.* (Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`) 6.1.2

Chen, M.-H., Shao, Q.-M., & Ibrahim, J. G. (2000). *Monte carlo methods in Bayesian computation*. Springer-Verlag. A

Chinchor, N. (1997). MUC-7 named entity task definition. In *Proceedings of the seventh message understanding conference.* 7.2.3

comScore. (2008, March). *YouTube.com accounted for 1 out of every 3 u.s. online videso viewed in january.* `http://www.comscore.com/press/release.asp?press=2111Please`. 1.2.5

Conover, W. J. (1971). *Practical nonparametric statistics*. John Wiley & Sons. 7.2.1.1, 2

Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. Wiley-Interscience. 5.1.1, 5.1.2.1, 5.2.2.2

Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines*. Cambridge University Press. 2.1, 6.1.2

Dave, K., Lawrence, S., & Pennock, D. M. (2003). Mining the peanut gallery: Opinion extraction and semantic classication of product reviews. In *Proceedings of the 12th international conference on world wide web* (pp. 519–528). 2.3

Del.icio.us. (n.d.). *Del.icio.us*. `http://del.icio.us/`. 4.5.3

DuBois, P. H. (1957). *Multivariate correlational analysis*. Harper. 7.2.3

Dudoit, S., Shaffer, J. P., & Boldrick, J. C. (2003, February). Multiple hypothesis testing in microarray experiments. *Statistical Science*, *18*(1), 71–103. 7.2.1.1

Efron, E. (1972). *The news twisters*. Manor Books. 1.2.5, 5.2.1

Elo, S. K. (1996). *PLUM: Contextualizing news for communities through augmentation*. Unpublished master's thesis, Massachusetts Institute of Technology. 1

Eugenio, B. D., & Glass, M. (2004, March). The kappa statistics: a second look. *Computational Linguistics*, *30*(1), 95–101. 7.2.3

Fass, D. (1991). met*: A method for discriminating metonymy and metaphor by computer. *Computational Linguistics*, *17*(1), 49–90. 8.1

Fisher, W., Doddington, G., & Goudie-Marshall, K. (1986). The DARPA speech recognition research database: specifications and status. In *Proceedings of the DARPA speech recognition workshop.* 7.2.3

Flickr. (n.d.). *Flickr.* `http://www.flickr.com/`. 4.5.3

Fortuna, B., Galleguillos, C., & Cristianini, N. (2008). Text mining: Theory and applications. In (chap. Detecting the bias in media with statistical learning methods). Taylor and Francis Publisher. 2.1

Google. (n.d.). *Google news.* `http://news.google.com/`. 1

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, *101*, 5228–5235. 2.5

Hauptmann, A., Yan, R., & Lin, W.-H. (2007). How many high-level concepts will fill the semantic gap in news video retrieval? In *Proceedings of the sixth international conference on image and video retrieval (civr).* (Runner-up of the Best Paper Aware in CIVR 2007) 5.2.3.4, 6.2.3.5

Hauptmann, A. G. (2004). Towards a large scale concept ontology for broadcast video. In *Proceedings of the third international conference on image and video retrieval (civr).* 5.2.1, 6.2.1

Hauptmann, A. G., Baron, R., Christel, M., Conescu, R., gao, J., Jin, Q., et al. (2005). CMU Informedia's TRECVID 2005 skirmishes. In *Proceedings of the 2005 TREC video retrieval evaluation.* 3.2.1, 5.2.3.4, 6.2.3.5

Hoew, J. (2006). The rise of crowdsourcing. *Wired Magazine*, *14*. 7.2.3

Hotelling, H. (1936, December). Relations between two sets of variates. *Biometrika*, *28*(3/4), 321–377. 2.1

Hovland, C. I., Janis, I. K., & Kelley, H. H. (1953). Communication and persuation. In (pp. 99–133). Yale University Press. 1

Ireson, B. (2004). "Minions". In *Proceedings of the twelfth ACM international conference on multimedia* (pp. 991–992). 2.6

Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. In *Proceedings of the 9th european conference on machine learning (ECML)*. 2.4, 6.1.2

Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., & Saul, L. K. (1999). An introduction to variational methods for graphical models. *Machine Learning*, *37*, 183–233. 4.2, 7.1.2

Kennedy, L., & Hauptmann, A. (2006, March). *LSCOM lexicon definitions and annotations (version 1.0)* (Tech. Rep. No. ADVENT 217-2006-3). Columbia University. 3.2.2, 5.2.2.1, 6.2.3.1

Kessler, B., Nunberg, G., & Schütze, H. (1997). Automatic detection of text genre. In *Proceedings of the 35th annual meeting on association for computational linguistics* (pp. 32–38). 2.4

Klimt, B., & Yang, Y. (2004). The enron corpus: A new dataset for email classification research. In *Proceedings of the 15th european conference on machine learning (ECML)* (pp. 217–226). 2.4

Kress, G., & Leeuwen, T. van. (1996). *Reading images: The grammar of visual design*. Routledge. 6.2.1

Krippendorff, K. (1980). *Content analysis*. Sage Publications. 1.3

Kull, S. (2003, October). *Misperceptions, the media and the iraq war.* 1.2.5

Kullback, S., & Leibler, R. A. (1951, March). On information and sufficiency. *The Annals of Mathematical Statistics*, *22*(1), 79–86. 5.1.1

Lakoff, G. (1992). Thirty years of linguistic evolution. In M. Pütz (Ed.), (chap. Metaphor and War: The Metaphor System Used to Justify War in the Gulf). Amsterdam: John Benjamins. 8.1

Lakoff, G. (2003, March 18). Metaphor and war, again. *AlterNet*. 8.1

Lewis, D. D. (1998). Naive (bayes) at forty: The independence assumption in information retrieval. In *Proceedings of the 9th european conference on machine learning (ECML)*. 2.2, 2.4, 5.1.1, 6.1.1

Lewis, D. D., Yang, Y., Rose, T. G., & Li, F. (2004). RCV1: A new benchmark collection for text categorization research. *Journal of Machine Learning Research*, *5*, 361–397. 2.4

Likert, R. (1932). *A technique for the measurement of attitudes* (Archives of Psychology No. 140). New York City: Columbia University. 7.2.3

Lin, W.-H., & Hauptmann, A. (2006). Do these documents convey different perspectives? a test of different perspectives based on statistical distribution divergence. In *Proceedings of the 44th conference on association for computational linguistics (acl).* 5.2.3.2

Lin, W.-H., & Hauptmann, A. (2008). Do these news videos portray a news event from different ideological perspectives? In *Proceedings of the second IEEE international conference on semantic computing.* 6.3

Lin, W.-H., Wilson, T., Wiebe, J., & Hauptmann, A. (2006). Which side are you on? identifying perspectives at the document and sentence levels. In *Proceedings of tenth conference on natural language learning (CoNLL).* 4.5.2, 7.2, 7.2.2.1

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press. 5.2.3.4, 6.2.3.2, 6.2.3.5

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. The MIT Press. 4.5.4, 5.1.1, 6.2.3.3

Many Eyes. (n.d.). *Using the tagcloud to compare two texts.* `http://services.alphaworks.ibm.com/manyeyes/page/Tag_Cloud.html`. 4.5.3

Mason, Z. J. (2004). Cormet: A computational, corpus-based conventional metaphor extraction system. *Computational Linguistics*, *30*(1), 23–44. 8.1

Mateas, M. (2002). *Interactive drama, art and artificial intelligence*. Unpublished doctoral dissertation, School of Computer Science, Computer Science Department, Carnegie Mellon University, Pittsburgh, PA 15213 USA. 2.6

Mateas, M., Vanouse, P., & Domike, S. (2000). Generation of ideologically-biased historical documentaries. In *Proceedings of the seventeenth national conference on artificial intelligence and twelfth conference on innovative applications of artificial intelligence.* 2.6

McCallum, A., Corrada-Emmanuel, A., & Wang, X. (2004). *The author-recipient-topic model for topic and role discovery in social networks: Experiments with Enron and academic email* (Tech. Rep. No. UM-CS-2004-096). Department of Computer Science, University of Massachusetts Amherst. 2.5

McCallum, A., & Nigam, K. (1998). A comparison of event models for naive bayes text classification. In *Proceedings of aaai-98 workshop on learning for text categorization.* 2.4

McDonald, R., Hannan, K., Neylon, T., Wells, M., & Reynar, J. (2007). Structured models for fine-to-coarse sentiment analysis. In *Proceedings of the 45th conference on association for computational linguistics (acl).* 7.1.1

Meyers, A., Ide, N., Denoyer, L., & Shinyama, Y. (2007). The shared corpora working group report. In *Proceedings of the linguistic annotation workshop.* 7.2.3

Morinaga, S., Yamanishi, K., Tateishi, K., & Fukushima, T. (2002). Mining product reputations on the web. In *Proceedings of the eigth ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 341–349). 2.3

Mosteller, F., & Wallace, D. L. (1984). *Applied bayesian and classical inference: The case of the federalist papers*. Springer-Verlag. 2.4

Mullen, T., & Collier, N. (2004). Sentiment analysis using support vector machines with diverse information sources. In *Proceedings of the 2004 conference on empirical methods in natural language processing.* 2.3

Nallapati, R., Ahmed, A., Xing, E. P., & Cohen, W. W. (2008). Joint latent topic models for text and citations. In *Proceedings of the 14th ACM SIGKDD international conference on knowledge discovery & data mining.* 8.1

Naphade, M., Smith, J. R., Tesic, J., Chang, S.-F., Hsu, W., Kennedy, L., et al. (2006, July–September). Large-scale concept ontology for multimedia. *IEEE Multimedia*, *13*(3), 86–91. 6.2.3.2

Naphade, M. R., & Smith, J. R. (2004). On the detection of semantic concepts at TRECVID. In *Proceedings of the twelfth ACM international conference on multimedia.* 5.2.3.4, 6.2.1, 6.2.3.5, 6.3

Nasukawa, T., & Yi, J. (2003). Sentiment analysis: Capturing favorability using natural language processing. In *Proceedings of the 2nd international conference on knowledge capture* (pp. 70–77). 2.3

Ng, A. Y., & Jordan, M. (2002). On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. In *Advances in neural information processing systems (NIPS).* 6.1.2

Ng, V., Dasgupta, S., & Arifin, S. M. N. (2006). Examining the role of linguistic knowledge sources in the automatic identification and classification of reviews. In *Proceedings of the 44th conference on association for computational linguistics (acl).* 2.3

Nigam, K., McCallum, A. K., Thrun, S., & Mitchell, T. (2000). Text classification from labeled and unlabeled documents using EM. *Machine Learning*, *39*(2/3), 103–134. 2.4

Ounis, I., Rijke, M. de, Macdonald, C., Mishne, G., & Soboroff, I. (2006). Overview of the trec-2006 blog track. In *Proceedings of the fifteenth text retrieval conference (trec 2006).* 2.2

Over, P., Ianeva, T., Kraaij, W., & Smeaton, A. F. (2005). TRECVID 2005 - an overview. In *Proceedings of the 2005 TREC video retrieval evaluation.* 3.2.1, 6.2.3.1

Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd conference on* *association for computational linguistics (acl).* 2.3

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, *2*(1–2), 1–135. 2.3

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of the acl-02 conference on empirical methods in natural language processing.* 2.3

Porter, M. (1980). An algorithm for suffix stripping. *Program*, *14*(3), 130-137. 3.1.1

R Development Core Team. (2005). *R: A language and environment for statistical computing.* Vienna, Austria. (ISBN 3-900051-07-0) A

Read, R. R., & Cressie, N. A. (1988). *Goodness-of-fit statistics for discrete multivariate data*. Springer-Verlag. 7.2.3

Riloff, E., & Wiebe, J. (2003). Learning extraction patterns for subjective expressions. In *Proceedings of the 2003 conference on empirical methods in natual language processing (EMNLP).* 2.2, 3.1.1

Riloff, E., Wiebe, J., & Wilson, T. (2003). Learning subjective nouns using extraction pattern bootstrapping. In *Proceedings of the seventh conference on natural language learning (CoNLL)* (pp. 25–32). 2.2

Ripley, B. D. (1987). *Stochastic simulation*. J Wiley. 5.1.1

Rivadeneira, A., Gruen, D. M., Muller, M. J., & Millen, D. R. (2007). Getting our head in the clouds: Toward evaluation studies of tagclouds. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 995–998). 4.5.3

Rosen-Zvi, M., Griffths, T., Steyvers, M., & Smyth, P. (2004). The author-topic model for authors and documents. In *Proceedings of the 20th conference on unvertainty in artificial intelligence.* 2.5

Sack, W. (1994). *Actor-role analysis: Ideology, point of view, and the news*. Unpublished master's thesis, Massachusetts Institute of Technology. 2.1, 2.1

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Lawrene Erlbaum Associates. 2.1

Sebastiani, F. (2002, March). Machine learning in automated text categorization. *ACM Computing Surveys*, *34*(1), 1–47. 2.4, 5.1.1

Shaffer, J. P. (1995). Multiple hypothesis testing. *Annual Review of Psychology*, *46*, 561–584. 7.2.1.1

Steinbach, M., Karypis george, & Kumar, V. (2000). A comparison of document clustering techniques. In *Proceedings of the 2000 KDD workshop on text mining.* 8.1

Steyvers, M., & Griffiths, T. (In Press). Latent semantic analysis: A road to meaning. In (chap. Probabilistic Topic Models). Laurence Erlbaum. 2.5

Sundheim, B. M. (1991). Overview of the third message understanding evaluation and conference. In *Proceedings of the third message understanding conference (MUC-3)*

(pp. 3–16). 2.1
Sunstein, C. R. (2007). *Republic.com 2.0*. Princeton University Press. 1, 1.2.5
techPresident. (2008, June). *YouTube stats*. http://www.techpresident.com/youtube. 1.2.5
Turney, P., & Littman, M. L. (2003). Measuring praise and criticism: Inference of semantic orientation from association. *ACM TOIS*, *21*(4), 315–346. 2.3
Uspensky, B. (1973). *A poetics of composition: The structure of the artistic text and typology of a compositional form*. University of California Press. (Translated by Valentina Zavarin and Susan Wittig) 2.2
Van Dijk, T. A. (1998). *Ideology: A multidisciplinary approach*. Sage Publications. 1, 1.1, 1.2.1, 1.2.5, 1.3, 2, 2.3, 4.5.3, 6, 6.3.2.1, 8.1
Wasserman, L. (2004). *All of statistics*. Springer. 5.1.2
Wiebe, J., Wilson, T., Bruce, R., Bell, M., & Martin, M. (2004). Learning subjective language. *Computational Linguistics*, *30*(3). 2.2, 2.2
Wiebe, J., Wilson, T., & Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, *39*(2–3), 165–210. 1, 2.2, 7.2.3
Wiebe, J. M. (1994). Tracking point of view in narrative. *Computational Linguistics*, *20*(2), 233–287. 2.2
Wilson, J. (1990). *Politically speaking: The pragmatic analysis of political language*. Blackwell. 8.1
Wu, C.-H. (2007). An empirical study on the transformation of likert-scale data to numerical scores. *Applied Mathematical Sciences*, *1*(58), 2851–2862. 7.2.3
Wu, X., Hauptmann, A. G., & Ngo, C.-W. (2007). Novelty detection for cross-lingual news stories with visual duplicates and speech transcripts. In *Proceedings of the 15th international conference on multimedia* (pp. 168–177). 2.6, 5.2.3.1
Xing, E. P. (2005, December). *On topic evolution* (Tech. Rep. No. CMU-CALD-05-115). Pittsburgh, PA: Center for Automated Learning & Discovery. 4.2, 7.1.2
Xing, E. P., Jordan, M. I., & Russell, S. (2003). A generalized mean field algorithm for variational inference in exponential families. In *Proceedings of the 19th annual conference on uncertainty in AI*. 4.2, 7.1.2
Yahoo. (n.d.). *My yahoo*. http://my.yahoo.com. 1
Yang, Y., & Liu, X. (1999). A re-examination of text categorization methods. In *Proceedings of the 22nd annual international ACM SIGIR conference on research and development in information retrieval* (pp. 42–49). 2.4
Yang, Y., & Pedersen, J. O. (1997). A comparative study on feature selection in text categorization. In *Proceedings of the fourteenth international conference on machine learning*. 2.4

Yu, H., & Hatzivassiloglou, V. (2003). Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the 2003 conference on empirical methods in natual language processing (EMNLP)* (pp. 129–136). 2.2, 2.3

Zhai, Y., & Shah, M. (2005). Tracking news stories across different sources. In *Proceedings of the 13th international conference on multimedia.* 2.6, 5.2.3.1

Zhang, D.-Q., Lin, C.-Y., Chang, S.-F., & Smith, J. R. (2004). Semantic video clustering across sources using bipartite spectral clustering. In *Proceedings of the 2004 IEEE international conference on multimedia and expo (ICME).* 2.6, 5.2.3.1

Zhou, C.-L., Yang, Y., & Huang, X.-X. (2007, March). Computational mechanisms for metaphor in languages: A survey. *Journal of Computer Science and Technology*, *22*(2), 308–319. 8.1