Journal of the Association for Information Science and Technology

Explore this journal >

View issue TOC
Volume 59, Issue 1
1 January 2008
Pages 98–110

Research Article

Ontology-supported polarity mining

Lina Zhou, Pimwadee Chaovalit

First published:
10 October 2007 Full publication history

DOI:
10.1002/asi.20735 View/save citation

Abstract

Polarity mining provides an in-depth analysis of semantic orientations of text information. Motivated by its success in the area of topic mining, we propose an ontology-supported polarity mining (OSPM) approach. The approach aims to enhance polarity mining with ontology by providing detailed topic-specific information. OSPM was evaluated in the movie review domain using both supervised and unsupervised techniques. Results revealed that OSPM outperformed the baseline method without ontology support. The findings of this study not only advance the state of polarity mining research but also shed light on future research directions.

Introduction

The ability to automatically extract and classify personal opinions from text would be enormously helpful to someone who sifts through vast amounts of news and Web data (Wilson, Wiebe, & Hwa, **2004**). Extracted opinions also can be used effectively by recommendation and collaboration systems.

Opinion mining aims to discover common patterns of user opinions from their textual statements automatically or semiautomatically. It is distinctively different from traditional text mining in that the latter is based on objective topics rather than on subjective perceptions. Specifically, traditional text mining focuses on specific topics (e.g., healthcare and travel) as well as topic shifts in text whereas opinion mining concerns the affective status of text regardless of its subject matter. Although there has been a recent surge of interest in opinion mining, the state-of-the-art techniques for opinion mining are much less mature than those for topic mining. This is partially attributed to the fact that topics are represented explicitly with keywords while opinions are expressed with subtlety. Opinion analysis requires deeper understanding of textual statements and thus is more challenging (Pang, Lee, & Vaithyanathan, **2002**). The focus of this study is polarity mining, a fundamental issue in opinion mining.

Polarity mining is a task of determining positive and negative orientations of textual information. The state of polarity mining has been significantly advanced by a host of previous studies (e.g., Bai, Padman, & Airoldi, **2005**; Das & Chen, **2001**; Dave, Lawrence, & Pennock, **2003**; Gamon, Aue, Corston-Oliver, & Ringger, **2005**; Hatzivassiloglou & McKeown, **1997**; Hu & Liu, **2004**; Morinaga, Yamanishi, Tateishi, & Fukushima, **2002**; Pang & Lee, **2004**; Pang et al., **2002**; Turney, **2002**; Turney & Littman, **2003**; Wilson, Wiebe, & Hoffmann, **2005**; Yi, Nasukawa, Bunescu, & Niblack, **2003**). Few of those studies incorporated topics into polarity mining. Although there is an emerging interest in the intersection of topic and polarity (Nigam & Hurst, **2004**), the focus is on whether a sentence is topical rather than on which specific aspects of a topic a sentence is about. The proposition of coupling polarity mining with entity extraction (Grefenstette, Qu, Shanahan, & Evans, **2004**) does not address the complexity of different aspects of a topic. A topic can get very complex by including a variety of subtopics, each describing a specific aspect of that topic. Moreover, the polarities of two subtopics of the same topic may be opposite to each other. Therefore, opinion mining may benefit from established techniques for topic mining. Motivated by the role of ontology in conceptualizing domain-specific information (Gruber, **1993**) and the potential of using multidimensional items in creating recommendations (Herlocker, Konstan, Terveen, & Riedl, **2004**), we propose ontology-supported polarity mining (OSPM). We expect that ontology can enhance polarity mining with domain-specific information. Moreover, it would enable informa-tion management at a finer and semantically self-contained granularity.

This article presents the first study on ontology-enhanced polarity mining. The results show that OSPM improved the performance of both supervised and unsupervised techniques for mining polarities of movie reviews. A movie review ontology developed in this study not only enhanced the performance of polarity mining but also improved our understanding and management of movie reviews. Some unique features of movie reviews identified in this study highlight some challenges of and provide new insights into polarity mining. Further, by examining the agreements between the overall polarity of a topic and the polarities of its subtopics, this study uniquely reveals that some properties of a domain are more important to shaping user's overall opinions than others. The findings of this study have significant practical implications to group decision support and customer relationship management. An increasing adoption of the Semantic Web will help realize the significance of OSPM.

The rest of the article is organized as follows. We first review concepts and techniques for polarity mining and ontology. Then, we introduce and empirically evaluate OSPM. Next, we discuss the findings of this study, their research and practical implications, and limitations and future directions.

Background

In this section, we first review the concepts and techniques related to polarity mining. Then, we describe ontology and discuss its potential in polarity mining.

Polarity Mining

Opinion mining is to extract and dissect opinions expressed in text reviews of a variety of genres of products and services such as automobiles, banks, movies, travel destinations (Turney, 2002), electronics (Dave et al., 2003; Morinaga et al., 2002), and mobile devices (Morinaga et al., 2002). Many Web sites such as Weblogs, bulletin boards, and news columns contain a large number of opinions, but lack a systematic organization of information based on such opinions. Some online review sites may not provide star or scale-based ratings and some users may misunderstand the meaning of ratings. As a result, the reviews from those Web sites and users would be difficult for other Web site visitors to use. Opinion mining, also called emotion mining, sentiment mining, attitude mining, or subjectivity mining, can be used to analyze product reputations (Morinaga et al., 2002) and to summa-rize consumer opinions into statistics information for search engines (Turney & Littman, 2003) and other applications.

In addition to polarity mining, there are two other types of opinion mining: subjectivity mining (Riloff & Wiebe, 2003; Yu & Hatzivassiloglou, 2003) and intensity mining (Wilson et al., 2004). Subjectivity mining is the process of separating opinion-oriented subjective statements and evaluations from objective statements expressing factual information. Intensity measures the strength of an opinion expressed in the text. For example, one negative review may be slightly more negative than another one. These different goals may intertwine in a specific opinion mining approach. For example, subjectivity mining can serve as a preprocessing step of polarity mining. We focus on polarity mining in the remainder of this article.

Polarity mining is a complex process. For example, raw data (e.g., reviews of products or services) are usually unstructured and heterogeneous. For example, product reviews exist in various forms across different Web sites (Dave et al., 2003), including e-commerce Web sites (e.g., Amazon.com), professional review sites (e.g., Cnet.com and ZDNet.com), consumer opinion sites (e.g., ConsumerReview.com and Epinions.com), news or magazine sites (e.g., Rollingstone. com), and bulletin boards or Usenet groups. These heterogeneous data need to be transformed into a consistent format.

The levels of text granularity on which to mine polarities need to be determined in advance. A large body of polarity mining research is conducted at the document level (e.g., Pang et al., 2002), some at the sentence level (e.g., Dave et al., 2003; Hatzivassiloglou & Wiebe, 2000; Nigam & Hurst, 2004), and only a few at the subsentence (e.g., clause) (Wilson et al., 2004), sentiment unit (Hiroshi, Tetsuya, & Hideo, 2004), or phrase level (Wilson et al., 2005). All of these studies are based on the structure of language units without considering domain knowledge. Polarities are also predicted at the level of reference (about a subject) or product attributes (Popescu & Etzioni, 2005; Yi et al., 2003), which are closest to our approach; however, none of them has exploited ontology in polarity mining. Additionally, they did not address how to derive the overall polarity from polarities of individual text segments.

Noisy raw data can further complicate polarity mining. For example, a Web page is likely to contain information other than reviews. Thus, it is necessary to separate reviews from nonreview data. Moreover, within a review, opinions are usually mixed with facts. Methods that have been proposed to address the aforementioned issues include objectivity classification (Finn, Kushmerick, & Smyth, 2002) and subjectivity analysis (Wiebe, 2000). Manual preparation becomes necessary when automatic methods do not warrant sufficiently accurate results.

Polarity mining is beneficial to consumers as well as business organizations. The results of polarity mining can help consumers decide which products to buy and help business organizations make appropriate strategic actions. Those benefits are empowered by machine learning techniques.

Polarity Mining Techniques

We summarize the extant techniques for polarity mining in Table 1. Based on whether they require training datasets, polarity mining techniques can be grouped into two broad categories: supervised and unsupervised. The supervised approach treats polarity mining as a classification problem. Reviews on Web sites (e.g., Cnet.com and Amazon) that provide star or scale-based ratings along with opinions serve as an ideal training set (Dave et al., **2003**) for supervised techniques. Feature selection is an important step in polarity mining. Some linguistic features that are commonly used to describe text include bag-of-words, n-grams (e.g., unigram, bi-grams, and tri-grams), word position, header information, and ordered word list (Mladenic, **1999**). Other features also have been incorporated into polarity mining, including semantic features based on substitutions and proximity (Dave et al., **2003**), frequent, noncontextual words in combination with various heuristics and annotators (Pang et al., **2002**), and a parsimonious vocabulary (Bai et al., **2005**) by capturing the dependencies among words. A classifier is trained on the feature values of training data, and then evaluated on test data to fine tune its performance.

Table 1. A summary of studies on polarity mining.

Studies	Polarity mining techniques used	Text granularity	Features	Data sources/Domains	Performance (accuracy)
Dave et al., **2003**	support vector machines	document	semantic features based on substitutions and proximity	Amazon Cnn.Net	88.9%
Pang et al., **2002**	Naïve Bayes, maximum entropy classification support vector machines	document	unigram, bi-gram, contextual effect of negation, feature presence or frequency, position	IMDb (Movie review)	82.9%
Pang & Lee, **2004**	Naïve Bayes support vector machines	document	sentence-level subjectivity summarization based on minimum cuts	IMDb	86.4%
Turney, **2002**	pointwise mutual information	document	bi-grams	Known positive terms such as excellent and negative terms such as poor movies, cars, banks	66–84%
Hatzivassiloglou & McKeown, **1997**	log-linear regression model	document	conjunctions part-of-speech	*Wall Street Journal corpus*	adjectives: precision: >90%
Das & Chen, **2001**	lexicons and grammar rules	document	words	financial news	62%
Bai et al., **2005**	two-stage Markov Blanket Classifier	document	dependence among words minimal vocabulary	IMDb Infonic	movie: 87.5% news: 89–96%

Studies	Polarity mining techniques used	Text granularity	Features	Data sources/Domains	Performance (accuracy)
Yi et al., 2003)	sentiment lexicon and semantic pattern	subject-spot terms	feature lexical semantics	digital camera music albums	85.6%
Morinaga et al., 2002	decision tree induction	document	characteristic words, co-occurrence words, and phrases	cellular phones, PDAs and Internet service providers	N/A
Turney & Littman, 2003	SO-LSA(Latent Semantic Analysis) SO-PMI (Pointwise Mutual Information) General Inquirer	document	words and phrases	TASA-ALLcorpus (from sources such as novels and newspaper articles)	65.27% (SO-LSA) 61.26% (SO-PMI)
Hu & Liu, 2004	Opinion word extraction and aggregation enhanced with WordNet	product features	opinion words opinion sentences;	Amazon Cnn.Net	digital camera 93.6% DVD player 73% MP3 player 84.2% cellular phone 76.4%
Chesley, Vincent, Xu, & Srihari, 2006	Support Vector Machines Wiktionary	document	textual features (e.g., exclamation points and question marks) and lexical semantics	Web sites of CNN, NPR, *Atlanta Journal and Constitution*, newspaper columns, reviews, political blogs …	positive: 84.2% negative: 80.3% objective: 72.4%
Gamon et al., 2005	Naïve Bayes classifier	sentence	stemmed terms, their frequency and weights, go list (salient words in a domain)	car review	recall: 96% (positive) 5˜24% (negative and other)
Popescu & Etzioni, 2005	relaxation labeling clustering	phrase	Syntactic dependency templates Conjunctions and disjunctions Morphological and WordNet	Amazon Cnn.Net	Opinion phrase polarity: precision: 86% recall: 97% relationships
Hiroshi et al., 2004	transfer-based machine translation principal patterns auxiliary/nominal patterns polarity lexicon	sentiment unit	full parsing semantic analysis	bulletin boards forums on digital cameras	precision: 89% recall 43%

Studies	Polarity mining techniques used	Text granularity	Features	Data sources/Domains	Performance (accuracy)
Nigam & Hurst, 2004	syntactic rules based chunking	sentence	a lexicon of polar phrases and their parts-of-speech syntactic patterns	online resources (e.g., Usenet, online message boards) in a particular domain	general polarity analysis: precision: 77% (positive), 84% (negative); recall: 43% (positive), 16% (negative)
Kennedy & Inkpen, 2005	support vector machines term-counting method a combination of the two	document		General Inquirer dictionary CTRW dictionary & Adj IMDb (Movie review)	enhanced combined method: 86.2%
Wilson et al., 2005	AdaBoost	phrase	subjectivity lexicon	multiperspective Question Answering Opinion Corpus	contextual polarity: 65.7%
Thomas, Pang, & Lee, 2006	support vector machines	speech segment	reference classification	2005 U.S. floor debate in the House of Representatives	with same-speaker links and agreement links: 71.16%

Many supervised machine learning techniques have been applied to polarity mining, including Support Vector Machines (Dave et al., 2003; Pang et al., 2002), Naïve Bayes (Gamon et al., 2005; Pang & Lee, 2004), maximum entropy classifier (Bai et al., 2005; Chaovalit & Zhou, 2005; Pang et al., 2002), AdaBoost (Wilson et al., 2005), Markov Blanket classifier, voted perceptron, and maximum entropy conditional random field learner (Bai et al., 2005). Bai et al. (2005) reported that standard machine learning techniques outperform human-produced baselines generated by counting the number of manually selected positive and negative indicator words in given documents. Additionally, a two-stage Markov Blanket Classifier, which has the ability to extract the most discriminating features for classification purposes, outperformed four other methods: Naïve Bayes, Support Vector Machine, voted perceptron, and maximum entropy conditional random field learner. The superior performance of the Markov Blanket classifier is attributed to the fact that it encodes and exploits conditional dependencies among words, which all other methods fail to capture.

An unsupervised approach to polarity mining generally relies on external knowledge resources beyond raw data and thus is knowledge-rich. In contrast with a supervised approach, an unsupervised approach does not involve an explicit training process in polarity mining. It generally proceeds in three steps:

1. Extracting words or phrases that express semantic orientations from text;
2. Determining the polarities of extracted words or phrases; and
3. Computing the polarity of the text by aggregating the polarities of individual words or phrases in the text.

External knowledge in support of unsupervised polarity mining comes in two broad types, semantic orientations (e.g., Hatzivassiloglou & Wiebe, 2000; Turney, 2002) and linguistic heuristics (e.g., Das & Chen, 2001; Tong, 2001; Yi et al., 2003), which are inspired by lexicography and cognitive linguistics, respectively. The semantic orientation of a word is measured by the relative difference in semantic scores between a target word and the average of all words in a text. For example, General Inquirer dictionaries (Buvac &

Stone, 2001) contain positive and negative categories, among others. Linguistic heuristic knowledge can be acquired from linguistic context and syntactic relations (Hatzivassiloglou & McKeown, 1997; Nigam & Hurst, 2004). The two types of knowledge complement each other and thus can be used together to support polarity mining. For example, the semantic orientations of words can be inferred from their linguistic associations with other words known to have positive or negative orientations (Hatzivassiloglou & McKeown, 1997; Turney & Littman, 2003), and contextual valance shifters even may change the original polarities of words (Kennedy & Inkpen, 2005).

In sum, in an unsupervised technique, associations with known orientations account for positivity and negativity whereas in a supervised technique, text features determine whether the text belongs to a positive or negative class. Unsupervised techniques are efficient and require little training; however, their performances are generally lower than those of supervised techniques in polarity mining. On the other hand, supervised models are subject to overtraining (Turney & Littman, 2003) and are highly dependent upon the quality and size of training data. In addition, it is time-consuming to create such models. Therefore, both techniques have strengths and weaknesses. In this study, we evaluated the proposed method with both types of techniques.

Ontology

Ontology, commonly referred to as the conceptualization of a domain (Gruber, 1993), aims to provide knowledge about specific domains that are understandable by both developers and computers. In particular, an ontology enumerates domain concepts and relationships among the concepts (Guarino, 1995), and provides a sound semantic ground of machine-understandable description of digital content. Ontology is popular in annotating documents with metadata, improving the performance of information retrieval and reasoning, and making data interoperable between different applications (Baziz, Boughanem, Aussenac-Gilles, & Chrisment, 2005; Duo, Juan-Zi, & Bin, 2005; Fensel, 2002; Khan, McLeod, & Hovy, 2004; Schreiber, Dubbeldam, Wielemaker, & Wielinga, 2001). In addition, an ontology-enabled semantic description of behaviors and services allows for better coordination of software agents in a multi-agent system (Hendler & McGuinness, 2001; Takeda, Iino, & Nishida, 1995; Takeda, Iwata, Takaai, Sawada, & Nishida, 1996). Therefore, ontology has a profound impact on a wide range of enterprise systems and information management applications.

Ontology appears especially promising for polarity mining. Many online products and services have primitive forms of ontologies, such as taxonomies available. An ontology allows us to interpret a text review at a finer granularity with shared meanings. Therefore, we propose to enhance polarity mining with ontology.

OSPM

Intuitively, polarity mining is a hybrid task falling somewhere between topic categorization and sentiment attribution. A human can easily detect the true sentiment of text, but bag-of-features classifiers would presumably find the task difficult since there are many words indicative of sentiment opposite to the whole text. A whole is not necessarily the sum of its parts (Turney, 2002). Fundamentally, some form of discourse analysis, or at least some way of determining the focus of each sentence, appears to be necessary so that one can decide whether a text is on topic. Moreover, a text usually focuses on only a few specific aspects of a topic. To address these issues, we propose OSPM. The use of ontology has the potential to refine and improve the process of polarity mining by identifying specific properties of a domain as well as relationships between different concepts from that domain.

Problem Formulation

Suppose $t \in T$ is a text about Domain d and T is a collection of texts. In addition, c is a concept in an ontology of d and is described with n properties: p_1, p_2, ..., p_n, about which sentiment orientations could be expressed in t. According to the number of related properties of c, t can be decomposed into m ($m \leq n$) segments.

As a result, the predicted polarity of t is not a single value but rather a vector [p_i, v_i, i = 1 ... n], consisting of polarity value $v_i \in [-1, 1]$ of property p_i of c. The values of −1, 0, and 1 indicate extremely negative, neutral, and extremely positive, respectively. The overall polarity value of t can be derived using Equation 1.

$$polarity(t) = \begin{cases} positive, & if\left(\frac{1}{m}\sum_{i=1}^{n} w_i v_i \geq 0\right); \\ negative, & otherwise. \end{cases} \quad (1)$$ (1)

where $w_i \in [0,1]$ denotes the weight of p_i.

An architecture for OSPM is illustrated in Figure 1. It extends the architecture for traditional polarity mining by incorporating the components of ontology development and ontology-based parsing. In the proposed architecture, polarity mining starts with collecting raw texts, most likely from the Internet. Preprocessing cleans up the raw data by removing noise that does not express content. Then, each text is parsed for extracting text segments and mapping text segments to the properties of an ontology concept. Finally, a polarity value is generated for each text segment in isolation and for the entire text as a whole. In view of the key role of ontology in OSPM, we devote the next section to ontology development.

Figure 1.

Open Figure

An architecture for ontology-supported polarity mining (OSPM).

Ontology Development

Ontology development can follow a bottom-up, a top-down, or a hybrid approach. The top-down approach starts with the high-level ontological concepts, which is then gradually expands into a full-fledged ontology. The bottom-up approach starts with textual documents and extracts ontological knowledge from the documents. A hybrid approach simultaneously derives knowledge from the top-level ontology and extracts low-level ontologies from documents, and then creates mappings between the different levels of ontologies. The hybrid approach has become increasingly popular in recent years because it is able to not only take advantage of existing domain resources (e.g., taxonomies) and heuristic knowledge but also discover new information from real documents. Despite some promising progress in ontology engineering, particularly in ontology learning, manual efforts remain indispensable in producing a usable ontology.

A hybrid approach was adopted to develop an ontology for movie reviews in this study. The top-down process of ontology development started with IMDb (www.imdb.com) metadata. IMDb houses a large collection of published movies. The information presented for each movie at IMDb is organized into five main sections: the title, production status, cast, crew, and miscellaneous. Miscellaneous entries consist of a number of subcategories such as awards, comments, country of origin, genre(s), keywords, languages, movie literature, filming location(s), running time, soundtrack information, and special effects company. They provide specific aspects of a movie which consumers would be interested in. For example, "cast" describes opinions about casting decisions. The bottom-up ontology development process involved analyzing the content of movie reviews, which will be introduced in the next section.

Content Analysis for Ontology Development

In this study, we carefully designed and performed content analysis to support ontology development. The process was guided by a preliminary ontology developed through the top-down process. The movie review concept was defined with properties, which were illustrated with both textual descriptions and sample instances.

A set of 180 movie review documents were randomly selected from an IMDb corpus for content analysis. Half of the movie reviews were positive and the other half negative. At the beginning, two human coders (A and B) were trained to annotate four separate movie reviews with the preliminary ontology. Each movie review was broken down into nonoverlapping segments, and each segment was assigned with exactly one property. Coder A then manually analyzed a subset of 110 reviews, and Coder B independently analyzed another subset of 130 reviews. There were 60 reviews in common between the two subsets for cross-validation. The two subsets and their overlapping sub-subsets all consisted of half positive and half negative movie reviews. Importantly, the coders were encouraged to create a new property if none of the existing properties of the movie review ontology covered any part of the review. Next, we compared the analysis results for the overlapping reviews between the two coders for interrater reliability. To this end, we developed Equation 2 to measure the reliability of encoding results R_{AB} by accounting for more than one property of an ontology concept related to each review text.

$$R_{AB} = \frac{1}{n}\sum_{i=1}^{n}\frac{|P_i^A \cap P_i^B|}{|P_i^A \cup P_i^B|} \qquad (2)$$

where P_i^A (or P_i^B) is the entire set of properties used to describe movie review i by Coder A (or B), and n is the total number of encoded movie reviews. The result showed that the reliability for positive and negative reviews were 78.3% and 89.1%, respectively. They were very encouraging considering that there were over 30 properties in total.

Any inconsistency between the two coders was resolved via discussion with the first author. Finally, the ontology was updated, and all review documents were resegmented and annotated using the updated ontology. Specifically, the following three types of updates were applied to the preliminary ontology: creating new properties (e.g., trailer, comparison, and implications), removing existing properties (e.g., date of release and production status), and refining existing properties (e.g., storytelling and writing). The final list of movie review properties is described in Table 2. The property *others* is a placeholder for other miscellaneous properties of movie reviews. *Attributes* and *storytelling* represent objective information rather than subjective opinions about a movie. Although the above two properties are always neutral in terms of semantic orientations, they were included in the final ontology to reflect the composition of an actual movie review.

Table 2. The descriptions of movie review properties.

Properties	**Description (Comments on … of a movie)**
action	specific action scene(s)

Properties	Description (Comments on … of a movie)
acting	the performance of actors and actresses
animation	computer-assisted motion
audience	target audience
cast	the cast in general, including casting decisions
characters	characters themselves and character development
cinematography	the art or technique of movie photography, including both the shooting and development
comedy	amusement, including funny, hilarious, and not funny
company	production company
comparison	a movie in comparison with another movie
costume	costume elements
directing	director(s) such as his/her working experience, idea, and past work
editing	the process of putting different scenes in sequence to tell a story
filmmaking	filmmaker(s) such as his/her working experience, idea, and past work
footage	length and quality
genre	genre(s) such as thriller, comedy, romantic, and drama
implications	social implications
others	aspects that are not covered by any other property
pacing	whether the pace is fast or slow
plot	plot and plot development
photography	nontechnical aspects of photographing such as visual effect of pictures
scene	specific scenes such as love scenes, and scenes in general
score	the music written by the original film composer
soundtrack	songs (either originally composed or transformed from recorded albums)
special_effect	optical or mechanical effects used to realize or create scenes
title	the title
trailer	the trailer

Properties	Description (Comments on … of a movie)
writing	the experience, history, and fame of script writer(s) as well as scripts or dialogue between characters
attributes	the content (e.g., director and cast)
storytelling	story or spoiler in movie critiques'jargon

Experiments

We conducted several experiments to empirically evaluate the performance of OSPM.

Data Selection

We selected movie review as the test domain for three reasons. First, movie review has been frequently used in previous polarity mining research (e.g., Pang et al., 2002). Movie review corpora that are publicly available provide good review quality, metadata, and reasonably large numbers of reviews and products (Dave et al., 2003). For example, extracted from IMDb archives of review newsgroups, the MovieLens datasets contain balanced positive and negative reviews of high quality, resulting from manual validation of sentiment polarities and manual cleaning of non-English and incomplete reviews. Second, movie is a common domain, and accordingly, a movie review ontology is easy for the general public to understand. Third, some special challenges associated with movie review mining present an opportunity for deepening our understanding of the problems and proposing new solutions. One problem is related to the change of general orientations of words when they are applied to the movie review context. For example, an "unpredictable" camera has a negative connotation, but a movie with an "unpredictable" plot indicates a positive orientation.

Data Preparation

We reused the movie reviews that were manually analyzed during ontology development in our experiment. The dataset consisted of 180 movie reviews, which were grouped into positive and negative categories. The dataset was prepared in two versions: original and segmented. The original version took the entire original reviews as units, which was treated as the baseline. Enabled by ontology support, the segmented version consisted of review segments that were generated based on the properties of the movie review ontology.

Polarity Mining Techniques

As reviewed earlier, both supervised and unsupervised techniques have been applied to polarity mining. They both can be used to implement the OSPM approach. For evaluation purposes, we selected one typical example of each type of the techniques to mine polarities in this study.

N-gram language modeling.

N-gram language modeling, the most widely used statistical language model, is a supervised machine learning technique. An n-gram language model is used to approximate the probability of a word sequence $S = w_1 w_2 \ldots w_N$ by assuming that w_i depends only on the preceding $(n - 1)$ words ($w_{i-n+1} \ldots w_2 w_1$) instead of all the previous words in S. This simplifies the model without losing much information because the dependency relationship between words in a text becomes weaker as their distance increases. The value of P(S), the occurrence probability of S, can be approximated with Equation 3:

$$P(S) = \prod_{i=1}^{N} P(w_i|w_{i-n+1} \cdots w_{i-1}) \quad (3)$$

When $n = 1$, no history is accounted for in calculating the probability of w_i. When $n = 2$, it becomes a bi-gram model, where the probability of w_i depends solely on the preceding word w_{i-1}. In general, the probabilities in Equation 3 can be estimated using Maximum Likelihood Estimation (MLE), as shown in Equation 4:

$$P(w_i \mid w_{i-n+1} \cdots w_{i-1}) = \frac{f(w_{i-n+1}, \ldots, w_i)}{f(w_{i-n+1}, \ldots, w_{i-1})}, \quad (4)$$

where $f(.)$ denotes the frequency of a substring in a text.

When language models are used for classification, the classification result is determined by the class label of the language model with the lowest perplexity/entropy value (Jelinek, **1990**).

The SRI Language Modeling Toolkit (Stolcke, **2002**) was used to train and test bi-gram language models for polarity mining. To address the data sparsity problem in building language models, smoothing methods have been proposed to reevaluate the zero-probability n-grams in MLE by discounting some probabilities from seen n-grams and assigning nonzero values to unseen n-grams. In addition, the back-off mechanism (Peng, Schuurmans, & Wang, **2004**) has been used to estimate the probability of an unseen gram from lower order grams. In view of the size of our datasets, we adopted bi-gram modeling enhanced with Kneser-Ney smoothing. The smoothing method was chosen due to its superior performance demonstrated in other studies. Further, vocabulary pruning was performed to boost the performance of language models.

For both original and segmented datasets, all the training data were used to develop language models. The models were then used to predict the polarity of each review in the test data. For the segmented version, the polarity of each review was determined as the majority of the predicted classes for all of its segments. In case of a tie, we randomly selected one of the two polarity classes.

General Inquirer.

We selected positive and negative categories from the General Inquirer to obtain their cumulative statistics for each review. The General Inquirer (Buvac & Stone, **2001**; Stone, Dunphy, Smith, & Ogilvie, **1966**) codes and assesses text features such as valence, emotion-laden words, Charles Osgood's three semantic dimensions, and cognitive orientations using a combination of the Harvard IV-4 dictionary, Harold Lasswell's dictionary, and social cognition work of Semin and Fiedler. There were 182 categories originally developed for social-science content-analysis research. Each category contains a list of words and word senses. The largest category is negative, consisting of 2,291 entries such as awkward, broke, and complexity. The positive category is slightly smaller, consisting of 1,915 entries such as admire, warm, and wise. Some language disambiguation routines such as stemming and word sense disambiguation were incorporated to improve the accuracy of category mapping.

For the original version of the dataset, we predicted the polarity of each review as the class with a higher statistics value. For the segmented version, the polarity of each review was determined as the majority of the predicted classes for all of its segments.

Evaluation Metrics

The performance of polarity mining was measured with accuracy, defined as the ratio of reviews that were classified correctly. Two other metrics—positive accuracy and negative accuracy—defined as the ratios of reviews that were classified correctly when positive reviews and negative reviews were considered separately. Also were adopted to identify possible bias of a polarity mining approach.

The language models were evaluated using 10-fold cross-validation to test the generality of the models on new data. Specifically, the data were randomly divided into 10 stratified samples. One of these samples was used to test the model that was trained using the remaining nine samples. Such a process was repeated for each of the 10 subsets. As an unsupervised technique, the General Inquirer was evaluated using the entire dataset. The results on the original datasets served as the baseline for comparison.

Results

The results of polarity mining are reported in Figure 2. The figure shows that OSPM increased the overall accuracy of the language modeling technique from 60.6 to 63.3%, and that of the General Inquirer technique from 63.9 to 72.2%. Moreover, compared with the baseline methods, OSPM tended to produce more balanced or less skewed results between positive and negative categories. Further, there was a significant increase of 7.1% in negative accuracy for the language modeling technique and an increase of 12% for the General Inquiry technique; however, there was a slight decrease in positive accuracy for both approaches (−2.2% for language modeling; −5.5% for the General Inquiry technique), which will be discussed in the next section.

Figure 2.

Open Figure

Results of polarity mining.

Discussion

The experiments demonstrated that ontology support had a positive impact on the performance of polarity mining, which is very encouraging. These findings have multifold theoretical contributions and practical implications to polarity mining.

Properties of the Movie Review Ontology

This study shows that the movie review ontology not only facilitates polarity mining at a finer granularity but also improves the performance of polarity mining. Additionally, the ontology provides an opportunity for discovering new heuristics for polarity mining.

Descriptive statistics of the movie review ontology properties are reported separately for positive and negative reviews in both Table 3 and Figure 3. It is shown that positive and negative reviews share the top-two properties: others and storytelling. These results highlight two facts about movie reviews that have implications for developing movie review mining approaches: Reviews tend to contain miscellaneous information and describe movie stories. The next nine most frequently used movie review properties (placed at 3–11 in overall ranking) were consistent between positive and negative reviews, including acting, attributes, cast, characters, comedy, directing, plot, scene, and writing, despite some slight differences in their relative rankings. Nonetheless, some other properties differed greatly in their relative rankings between positive and negative reviews. For example, genre, score, and costume were included in 11, 4, and 3 positive movie reviews, respectively, but they were never a part of negative reviews. Additionally, positive reviews were more likely to involve movie soundtrack and cinematography than were negative reviews; whereas negative reviews were more likely to engage audience and film making than were positive reviews. This information can be used as heuristics for predicting the polarity of movie reviews, and also may benefit movie production companies by allowing them to focus on improving those movie properties about which consumers care most.

Figure 3.

Open Figure

Property frequency comparisons between positive and negative reviews.

Table 3. Descriptive statistics of movie review properties.

Negative reviews				Positive reviews			
Order	**Properties**	**Frequencies**	**%**	**Order**	**Properties**	**Frequencies**	**%**
1	storytelling	85	15.2	1	others	84	14.3
2	others	82	14.6	2	storytelling	84	14.3
3	plot	55	9.8	3	acting	59	10.0
4	acting	51	9.1	4	attributes	43	7.3
5	directing	39	7.0	5	characters	42	7.2
6	attributes	38	6.8	6	plot	41	7.0
7	characters	35	6.3	7	directing	37	6.3
8	writing	33	5.9	8	writing	27	4.6
9	comedy	31	5.5	9	scene	26	4.4
10	scene	26	4.6	10	comedy	22	3.7
11	cast	19	3.4	11	cast	17	2.9
12	filmmaking	10	1.8	12	special_effect	13	2.2
13	special_effect	9	1.6	13	genre	11	1.9
14	action	7	1.3	14	action	9	1.5

Negative reviews				Positive reviews			
Order	Properties	Frequencies	%	Order	Properties	Frequencies	%
15	audience	7	1.3	15	photography	9	1.5
16	pacing	7	1.3	16	soundtrack	9	1.5
17	photography	6	1.1	17	pacing	7	1.2
18	company	5	0.9	18	comparison	6	1.0
19	comparison	3	0.5	19	filmmaking	6	1.0
20	implication	3	0.5	20	cinematography	5	0.9
21	footage	2	0.4	21	footage	5	0.9
22	soundtrack	2	0.4	22	editing	4	0.7
23	title	2	0.4	23	implication	4	0.7
24	cinematography	2	0.4	24	score	4	0.7
25	editing	1	0.2	25	audience	3	0.5
26	trailer	1	0.2	26	costume	3	0.5
27	animation	1	0.2	27	trailer	3	0.5
28	costume	0	0.0	28	animation	1	0.2
29	genre	0	0.0	29	company	1	0.2
30	title	0	0.0	30	title	1	0.2

Contradicting Polarities between Parts and the Whole

An interesting observation from our experiments was that an overall positive rating for a movie did not entail positive ratings for all of its properties. Conversely, despite an overall negative rating, a movie could receive positive comments on some of its properties. Therefore, we examine the relationship between the polarities of individual properties and the polarity of entire movie reviews. To this end, we developed a metric called *opposition* to measure the strength of a negative association. Adapted from support metric used in association rule mining, opposition was defined as the percentage of property occurrences that have opposite polarities to the overtone of the entire reviews.

Table 4 shows that some properties are more likely to express semantic orientations opposite to those of entire reviews than are others. For example, the opinions on pacing in positive reviews were mostly negative, and the opinions on special effect in negative reviews were mostly positive. Some other properties such as comedy and soundtrack also received high opposition scores, as shown in Table 4. These findings would help us refine polarity mining techniques by adjusting the weights of individual properties according to their opposition scores. For example, the weights of those properties with high opposition scores should be discounted in mining the polarity of movie reviews.

Table 4. Opposition statistics of movie review properties.

Negative reviews		Positive reviews	
Properties	**Opposition (%)**	**Properties**	**Opposition (%)**
special_effect	55.6	Pacing	71.4
soundtrack	50.0	comedy	31.8
acting	41.2	others	31.0
cast	36.8	writing	29.6
implications	33.3	character	28.6
comedy	19.4	plot	26.8
photography	16.7	editing	25.0
directing	15.4	implications	25.0
scene	15.4	special_effect	23.1
audience	14.3	acting	20.3
action	14.3	footage	20.0
character	11.4	cast	17.6
others	11.0	directing	16.2
filmmaking	10.0	scene	15.4
plot	7.3	photography	11.1
writing	6.1	soundtrack	11.1
storytelling	1.2	genre	9.1

Challenges of Movie Review Mining and Suggested Solutions

Movie reviews pose a number of challenges to polarity mining. The challenges can be partly explained with several unique features of movie reviews, which help suggest possible solutions. First, there is a wide range of word choices for writing reviews, which introduces the sparsity problem to building language models using bag-of-words. The aforementioned problem could be partly alleviated by increasing the data size and by incorporating synonyms. In addition, stop words were found to make frequent appearances in bi-gram features and to dominate the polarity mining models in our experiments. Thus, it is helpful to develop a mechanism for stop-word filtering while building language models. For example, we only kept bi-grams containing at least one nonstop word (e.g., carry on).

Second, movie reviews are infused with factual information (e.g., storytelling), which could bias polarity prediction results. For example, some movie names such as “main failure,” “cheap looking,” “beautiful day,” and “decent film” contain words indicating semantic orientations, which may introduce ambiguities to polarity mining. Additionally, it is not unusual for a good movie to contain violent scenes and unhappy endings, and convey tragic stories. For example, in a positive review of the movie *Cabinet of Dr. Caligari*, “horrific” and “insane” have negative connotations.

Subjectivity analysis (Wiebe, **2000**; Yu & Hatzivassiloglou, **2003**) is a promising way to separate opinions from facts. Domain resources could be incorporated to resolve ambiguous words in a review, which is discussed in detail in our future work.

Third, the ironic tone expressed with sarcastic words in some reviews is prone to misinterpretation. It may require deep linguistic analysis such as discourse analysis for the detection of a sarcastic review style. For example, in the following excerpt from a negative review of the movie *About Schmidt*, “terrifically” and “very well” express positive emotions.

“As with most highly depressing movies, the story in *About Schmidt* dragged very slowly for the first 100% of it (The last 0% was actually very well paced and terrifically written).”

Fourth, positive movie reviews tend to contain similar numbers of properties with positive and negative orientations, but the positive properties are expressed in a more intense manner than are the negative ones. As a result, the performance of OSPM in terms of positive accuracy was reduced, as shown in Figure 2. This indicates the need to look into the intensity of semantic orientations in polarity mining.

Practical Implications

OSPM offers multifold practical benefits. First, it saves time for e-commerce consumers by allowing them to focus on specific aspects of products and services to their inter-ests instead of reading through numerous texts in their entirety. Ultimately, OSPM would help consumers make better purchasing decisions. Second, it improves information presentation by constructing multiperspective information extraction and summarization. For example, OSPM could be integrated into search engines to provide quick statistics of search results such as “5,000 hits found on Web 2.0, 80% of which are positive about social impact, and 20% are negative about privacy impact.” It would enable information management at a fine and semantically self-contained granularity. Third, OSPM can be used to summarize user input and feedback (Pang et al., **2002**), debating results (Thomas, Pang, & Lee, **2006**), and group opinions. Fourth, it can improve the performance of personalization and recommendation systems by selectively presenting the information according to users' preferences and interests.

All of the above-mentioned implications can lead to improved customer relationships in businesses and organizations, which is a major driving force behind polarity mining. For example, consumer complaints would be addressed in a timely manner, and recommendations would be made as a more informed decision to improve consumer trust and loyalty. OSPM allows businesses to prioritize various properties of their products and services according to their influences on shaping consumers' positive or negative attitude and to increase their return on investment. Additionally, OSPM would help governments and organizations keep track of the public's attitudes and feelings about policies or events in news and other online media.

Limitations and Future Work

Ontology development and review parsing currently rely on manual resources. The experiments reported here explored whether ontology support was useful for the automatic prediction of polarity, which was clearly proven. The next step will be the automatic creation of such resources. Ontology learning is a process of learning an ontology automatically from related textual documents and resources (Zhou, 2007). In addition, subjectivity analysis can be conducted to differentiate subjective expressions from objective facts using some lexical resources (Wiebe, 2000). Further, text decomposition in accordance with semantics-based text themes (Salton, Singhal, Buckley, & Mitra, 1996) is an important future task. In particular, information extraction (Srihari, Li, Cornel, & Niu, 2006), feature identification (Popescu & Etzioni, 2005), aspect identification (Kobayashi, Iida, Inui, & Matsumoto, 2006), or discourse analysis (Teufel & Moens, 2002) facilitate the identification of text segments related to specific properties of an ontology concept. For example, 13 sentential features were extracted to classify scientific articles into a fixed set of rhetorical categories (Teufel & Moens, 2002).

This study highlights the subtlety, ambiguity, and diversity of natural languages that people use in expressing their opinions. There clearly is a need to further refine features indicating semantic polarities for unsupervised techniques. First, generic semantic lexicons can be customized for specific domains. In addition, domain lexicons or dictionaries may contribute to disambiguation. For example, databases of movie names, actors, and directors could be employed to reduce noise in the data while mining the polarities of movie reviews. Further, given that the contextual polarity of a phrase in which a word appears may be different from the word's prior polarity (Wilson et al., 2005), contextual polarity identification can be incorporated to improve the accuracy of polarity classification. For example, contextual valence shifters (Kennedy & Inkpen, 2005; Polanyi & Zaenen, 2006), including negations, intensifiers, and diminishers, can be used to change the polarity or the degree of polarity of a particular term.

Regarding the performance of supervised polarity mining, improvements can be expected by using more representative or larger training sets. Testing the generality of OSPM in other domains will also be interesting. In our experiments with language modeling techniques, we built a single language model across all the properties contained in the training data in each fold rather than a separate language model for each property. As a result, the language models trained for OSPM were close to those for the baseline method despite their different performances on test datasets. We believe that the performance of OSPM using language modeling techniques would be improved further by developing a separate language model for each movie review property. Additionally, the findings of this study show that different movie review properties are not equally important to polarity mining, and positive comments are likely to be expressed with a higher level of intensity than are negative comments. Thus, property selection and weighting would further improve the accuracy of polarity mining. Furthermore, based on the discourse patterns of movie reviews, we may create heuristics to support polarity mining by focusing on a subset of discourse elements. For example, we observed that a movie review tends to begin with an expression setting the overall tone of the review, proceed with a plot discussion, and end with a summary of viewpoints.

Each movie review segment was assigned with one and only one property in this study. For ambiguous cases, only the predominant type of property was chosen. In future studies, multiple-property assignments with varying degrees of membership may be considered. In cases of informal text that do not follow the convention of punctuation usage, we recommend adopting subsentence units in ontology-based parsing.

Conclusion

Polarity mining is a challenging sentiment classification problem. The proposed OSPM approach and related findings in the movie review domain not only advance the state of polarity mining research but also shed light on several promising future directions. The increased availability of domain ontologies has paved the way for the wide adoption of OSPM in practice.

Acknowledgment

The authors thank Yongmei Shi for her help in developing language models for polarity mining. This work is supported in part by a DRIF fund from UMBC.