1-1-2009

Automatic Domain Ontology Extraction for Context-Sensitive Opinion Mining

Raymond Y.K. Lau
City University of Hong Kong, raylau@cityu.edu.hk

Chapmann C.L. Lai
City University of Hong Kong, chunllai@cityu.edu.hk

Jian Ma
City University of Hong Kong, isjian@cityu.edu.hk

Yuefeng Li
Queensland University of Technology, Y2@qut.edu.au

AUTOMATIC DOMAIN ONTOLOGY EXTRACTION FOR CONTEXT-SENSITIVE OPINION MINING

Completed Research Paper

Raymond Y.K. Lau
Department of Information Systems
City University of Hong Kong
Tat Chee Avenue, Kowloon,
Hong Kong SAR
raylau@cityu.edu.hk

Chapmann C.L. Lai
Department of Information Systems
City University of Hong Kong
Tat Chee Avenue, Kowloon,
Hong Kong SAR
chunllai@cityu.edu.hk

Jian Ma
Department of Information Systems
City University of Hong Kong
Tat Chee Avenue, Kowloon,
Hong Kong SAR
isjian@cityu.edu.hk

Yuefeng Li
School of Information Technology
Queensland University of Technology
GPO Box 2434, Brisbane, Qld 4001
Australia
Y2@qut.edu.au

Abstract

Automated analysis of the sentiments presented in online consumer feedbacks can facilitate both organizations' business strategy development and individual consumers' comparison shopping. Nevertheless, existing opinion mining methods either adopt a context-free sentiment classification approach or rely on a large number of manually annotated training examples to perform context-sensitive sentiment classification. Guided by the design science research methodology, we illustrate the design, development, and evaluation of a novel fuzzy domain ontology based context-sensitive opinion mining system. Our novel ontology extraction mechanism underpinned by a variant of Kullback-Leibler divergence can automatically acquire contextual sentiment knowledge across various product domains to improve the sentiment analysis processes. Evaluated based on a benchmark dataset and real consumer reviews collected from Amazon.com, our system shows remarkable performance improvement over the context-free baseline.

Keywords: Opinion Mining, Sentiment Analysis, Kullback-Leibler divergence, Fuzzy Sets, Domain Ontology, Ontology Extraction, Sentiment Context.

Introduction

With the norm of users contributed data in the era of Web 2.0, increasingly more people have submitted or retrieved individual viewpoints about products, organizations, or political issues via a variety of Web-based channels such as Blogs, forums, social networks, and e-Commerce sites. Due to the problem of information overload (Lau et. al. 2008; Lau and Lai 2008), manually browsing a large number of consumer reviews posted to the Web may not be feasible, if not totally impossible. The huge volume of documents (e.g., consumer reviews) archived on the Web has triggered the development of intelligent tools to automatically extract, analyze, and summarize their contents. Opinion mining is also referred to as opinion analysis, sentiment analysis, or subjectivity analysis (Abbasi et. al. 2008; Turney and Littman 2003; Wright 2009). Opinion mining differs from Information Retrieval (IR) in that it aims at extracting the viewpoints about some entities rather than simply identifying the topical information about the entities (Macdonald and Ounis 2007; Wilson et. al. 2004). Analyzing the sentiment of consumer feedbacks posted to Blogs, forums, or e-Commerce sites can generate huge business values for organizations (Archak et. al. 2007; Danescu-Niculescu-Mizil et. al. 2009). Although consumer reviews are subjective in nature, these reviews are often

considered more creditable and trustworthy than other traditional information sources from the perspectives of customers (Bickart and Schindler 2001; Wright 2009). In this paper, we will illustrate a novel opinion mining methodology which can automatically extract, analyze, and summarize consumers' reviews about various products with reference to the specific product contexts.

The Research Challenges and Our Contributions

Though traditional sentiment analysis or opinion mining was performed at the document level (Dave et. al. 2003; Turney and Littman 2003), increasingly more research has examined opinion mining at the more fine-grained sentence or phrase level in recent years (Agarwal et. al. 2009; Xu et. al. 2008; Wilson et. al. 2005). Even if a review (i.e., document) is rated as positive, negative sentiments could appear in the same review. Therefore, opinion mining against consumers' reviews is often performed at the product feature level to provide deep analytics for the target product (Archak et. al. 2007; Hu and Liu 2004; Popescu and Etzioni 2005). The quest for a more fine-grained opinion mining method is driven by the fact that sentiment words are often context-dependent (Agarwal et. al. 2009). For instance, while the token "small" in the expression "the hotel room is so small" implies a negative sentiment, the same token may have a positive meaning in another situation such as "it's so convenient to bring a small notebook for a business trip". Another example is that "unpredictable" has a negative orientation in the context of "automotive". However, the same sentiment has a positive orientation such as "unpredictable plot" in the context of "movie". In fact, the token "unpredictable" has a strong negative orientation defined in sentiment lexicons such as OpinionFinder (Wilson et. al. 2005) and SentiWordNet (Esuli and Sebastiani 2005). Therefore, using lexicon-based approach alone may not provide an effective solution for context-sensitive opinion mining.

Linguistic or inference-based method can deal with sentiment analysis for some general cases, but there are many situations (particularly down to the phrase level) that the general rules or inference process could not be applied. For example, no general linguistic rule can be applied to detect the polarity of the sentiment "small" in the expression "The camera is good in general; the viewer panel is small". On the other hand, machine learning methods usually require a large number of manually labeled training examples to build an accurate classifier. Nevertheless, manually annotating a large number of review messages at the phrase level is extremely labor intensive and expensive. Even though attempts are made to mine consumers' reviews at the product feature level, the polarities of sentiments are assumed the same across product domains (i.e., context-free) (Archak et. al. 2007; Hu and Liu 2004; Popescu and Etzioni 2005). For instance, "small" is often assumed negative no matter it is referring to a hotel room or the physical size of a Netbook computer. Indeed, it has been pointed out that developing an automated technique for building sentiment lexicon is an important topic for research and practices in opinion mining (Macdonald and Ounis 2007), and contextual domain knowledge is essential to improve the performance of opinion mining systems (Bao et. al. 2008)

The main contributions of our research are: (1) the design of a novel context-sensitive opinion mining methodology to improve the effectiveness of sentiment analysis; (2) the extension of a fuzzy domain ontology extraction method (i.e., learning the non-taxonomic fuzzy relations) for the automatic construction of sentiment lexicons; (3) the development of a novel computational method to predict the context-sensitive polarities of sentiments; (4) the design and development of a prototype system for context-sensitive sentiment analysis. The practical implication of our work is that an effective opinion mining methodology is developed to enhance both organizations' business strategy development and individuals' comparison shopping processes.

Research Methodology

Our research work is driven by the "Design Science" research methodology (Hevner et al. 2004). The design science research methodology emphasizes on the discovery of novel knowledge of a problem domain by the construction and application of "designed artifacts". Such artifacts should also be rigorously evaluated, and they should contribute to address relevant business problems. For our research, the designed artifacts include a methodology for context-sensitive opinion extraction and prediction, a fuzzy domain ontology based computational model for the representation of context-sensitive sentiment lexicon, and an instantiation of the design by the construction of a Web-based context-sensitive opinion mining prototype system. Our design is based on sound theories with rigorous theoretical foundations. For example, the fuzzy domain ontology extraction method is developed based on fuzzy sets and fuzzy relations (Zadeh 1965) which offer the expressive power to capture the uncertainty presented in opinion mining. Our approach of predicting the polarities of sentiments is based on a well-known statistical learning technique, a variant of Kullback-Leibler divergence (Kullback and Leibler 1951).

Moreover, our designed artifacts are rigorously evaluated based on a benchmark e-Commerce dataset and real consumer reviews retrieved from a popular e-Commerce Website. Above all, our designed artifacts make significant contributions to improve both organizations and individuals' capabilities of analyzing the sheer volume of consumer feedbacks posted to the Web these days. As a result, organizations can develop appropriate marketing and product design strategies quickly and individuals can conduct comparison shopping easily. As a whole, our research is driven by the processes of designing and developing the artifacts (e.g., a fuzzy domain ontology based computational model for context-sensitive opinion mining) to improve business strategy development and individuals' online shopping experience. The main research questions can be summarized as follows:

How can we apply a fuzzy domain ontology extraction method to automatically build domain specific sentiment lexicons to facilitate context-sensitive opinion mining?

Can we develop an effective sentiment polarity classification method which does not rely on extra human effort to annotate training examples?

Is the proposed ontology-based context-sensitive opinion mining approach more effective than a context-free opinion mining approach?

Outline of the Paper

The remainder of the paper is organized as follows. The next section highlights previous research related to opinion mining and ontology learning, which is followed by the architectural design of an ontology-based context-sensitive opinion mining system. The computational models for fuzzy domain ontology extraction and context-sensitive opinion mining are then illustrated. The quantitative evaluation of our prototype system is reported afterwards. Finally, we offer concluding remarks and describe future direction of our research work.

Related Research

A light weight fuzzy domain ontology extraction method has been developed to automatically generate concept hierarchies based on textual contents extracted from online message boards (Lau et. al. 2009). The algorithm of fuzzy domain ontology extraction includes concept extraction, concept pruning, dimensionality reduction, and fuzzy relation extraction. Fuzzy relation extraction involves the generation of taxonomic relations using the structural similarity (SSIM) metric developed in the field of image analysis. Formal concept analysis (Cimiano et. al. 2005) and fuzzy formal concept analysis (Tho et. al. 2006) have also been applied to build domain ontology automatically. Formal concept analysis is a systematic method for deriving implicit relationships among concepts described by a set of attributes (Wille 2005). For the research work reported in this paper, we utilize a simplified version of the fuzzy domain ontology model (Lau et. al. 2009) for sentiment knowledge representation. In particular, we develop effective computational methods to learn the non-taxonomic relations among concepts (e.g., products, product features, and sentiments) to support context-sensitive opinion mining.

An econometric opinion mining method has been proposed to analyze product feature evaluations expressed in online consumer reviews (Archak et. al. 2007). Each product feature is represented by a noun which frequently appears in the consumer reviews. A manual procedure is then involved to filter the candidate nouns to identify correct product features. The adjectives collocated with product features are taken as the sentiment words. A pair of product feature and sentiment (also called an opinion phrase) is formally represented by a vector in the tensor product space. Hedonic regressions are applied to estimate the relative weights of product features and the strength of the sentiments associated with those features. OPINE employs the "relaxation labeling" classification method developed by the computer visioning research community to detect sentiment polarity (Popescu and Etzioni 2005). Similarly, Feature-Based Summarization (FBS) system has been developed to extract explicit product features and sentiments at the sentence level (Hu and Liu 2004). The Apriori association rule mining algorithm is applied to extract the product features (i.e., noun phrases) frequently occurring in product reviews. A similar product feature extraction method is also applied to a product review mining system (Miao et. al. 2008). The ReviewSeer system adopts an *n*-gram approach for feature extraction and a machine learning approach for sentiment polarity classification (Dave et. al. 2003). For the aforementioned opinion mining systems, polarity detection of sentiments is not conducted with respect to a particular product domain. Our proposed opinion mining approach supports context-sensitive polarity detection rather than assuming that the polarity of a sentiment is the same across different product domains.

A hybrid lexicon and machine learning based approach has been applied to extract the sentiments from online stock message boards and then classify the discussions as bullish, bearish, or neutral (Das and Chen 2007). Sentiment identification is conducted based on the General Inquirer sentiment lexicon (Stone et. al. 1966); five statistical or machine learning classifiers coupled with a voting scheme are applied to classify the polarity of each message. Our approach differs in the sense that we focus on applying a statistical learning method to automatically build a context-sensitive sentiment lexicon rather than relying on the manually crafted sentiment lexicons for polarity detection. Entropy Weighted Genetic Algorithm (EWGA) has been developed to select the best syntactic (e.g., POS pattern) and stylistic features (e.g., number of special characters used in a document) for multilingual (e.g., English and Arabic) sentiment classification against various extremist online forums (Abbasi et. al. 2008). The EWGA algorithm selects the most informative features (e.g., n-gram[1]) according to information gain and passing those features to a SVM classifier for polarity classification (e.g., positive or negative) at the document level. Based on the technique of bootstrapping, a classification accuracy of 91% is achieved over a benchmark movie dataset (Pang et. al. 2002). Instead of employing machine learning approaches, our opinion mining system utilizes a statistical learning method i.e., a variant of Kullback-Leibler divergence, to detect the polarity of sentiments at the product feature level.

In the field of IR, Probabilistic Latent Semantic Analysis (PLSA) which is underpinned by the unigram language modeling approach is proposed to predict sentiment orientations in movie blog posts (Liu et. al. 2007). The PLSA model is combined with a time series analysis model (called autoregressive model) to predict the gross revenues of movies. PLSA is also applied to combine opinions expressed in a well-written expert review with those retrieved from Web 2.0 sources such as blog posts to generate a comprehensive opinion summary about a product or a political figure (Lu and Zhai 2008). Probabilistic generation language models are explored to identify and rank sentiment expressions at the document level (Zhang and Ye 2008). Instead of applying a probabilistic language modeling approach to opinion mining, we propose to address the problem of opinion mining using a fuzzy approach e.g., modeling the association between a product feature and a sentiment in terms of a fuzzy relation. In the field of machine learning, the problem of automatically identifying sentiment orientations across different domains is called the "Domain-Transfer" problem (Tan et. al. 2007; Tan et. al. 2008). A method called Relative Similarity Ranking (RSR) is proposed to select the most informative unlabeled opinionated documents from a training set to re-train a classifier (e.g., Support Vector Machine). Instead of identifying the most informative training examples, we employ an efficient statistical learning technique to automatically build a domain dependent sentiment lexicon based on the training set pertaining to each product domain.

Linguistic rules are applied to detect the context-sensitive orientations of sentiments extracted from online customer reviews (Ding and Liu 2007). For example, for the sentence "This camera takes great pictures and has a long battery life", the orientation of the sentiment "long" is classified as positive because it is conjoined with the positive seeding sentiment "great". An inference-based opinion mining method called Semantic Orientation (SO) analysis has been developed to estimate the polarity of sentiments (Hatzivassiloglou and McKeown 1997; Turney and Littman 2003). The SO of an arbitrary word can be estimated based on the strength of association between the word and fourteen seeding sentiment words such as good, nice, bad, poor, and so on. Point-wise Mutual Information (PMI) is proposed to compute the strength of association between any pair of words. Our system also employs a variant of Mutual Information to estimate the strength of associations between product features and sentiment words. However, polarity detection is underpinned by a variant of Kullback-Leibler divergence.

Context-sensitive sentiment analysis has been an active research topic in the Natural Language Processing (NLP) research community (Wilson et. al. 2005; Wilson et. al. 2006). A sentence is first parsed and represented by a dependency tree. A set of linguistic features are used to train the AdaBoost classifier to predict the sentiment orientation of a target word. An appraisal group is represented by a set of attribute values in some task-independent semantic taxonomies such as attitude, orientation, graduation, and polarity (Whitelaw et. al. 2005). The appraisal group method has been applied to analyze the sentiments of a movie review corpus. Apart from utilizing the fuzzy domain ontology, our system also employs basic syntactical features to infer sentiment polarity. However, instead of using sophisticated NLP techniques which are computationally expensive, we adopt a light-weight NLP approach so that our opinion mining system can scale up for the sheer volume of users contributed feedback data generated in the era of Web 2.0.

---

[1] An n-gram is a term with n consecutive words.

The Architectural Design of an Ontology Based Opinion Mining System

The general system architecture of our Ontology Based Product Review Miner (OBPRM) system is depicted in Figure 1. A user first selects a product category and a specific product for opinion mining (Task 1 in Figure 1). Based on the selected target product, the OBPRM system will use the Web services or APIs provided by e-Commerce sites (e.g., Amazon.com[2] and Cnet.com[3]) and Internet Search Engines (e.g., Google[4]) to retrieve the consumer reviews for the particular product (Task 2 in Figure 1). In addition, the crawlers of our system can also be invoked to retrieve information about product features and download consumer reviews (Task 3 in Figure 1). Traditional document pre-processing procedures (Salton et. al. 1975; Salton and McGill 1983) such as stop word removal, Part-of-Speech (POS) tagging, and stemming (Porter 1980) are then invoked to process the consumer reviews and product descriptions (Task 4 in Figure 1). We develop our POS tagger based on the WordNet lexicon (Miller et. al. 1990) and the publicly available WordNet API[5]. Similar to previous studies, a product feature is represented by a Noun or a Noun compound (Archak et. al. 2007; Hu and Liu 2004; Popescu and Etzioni 2005), and sentiment words are represented by Adjective or Adverb (Subrahmanian and Reforgiato 2008).
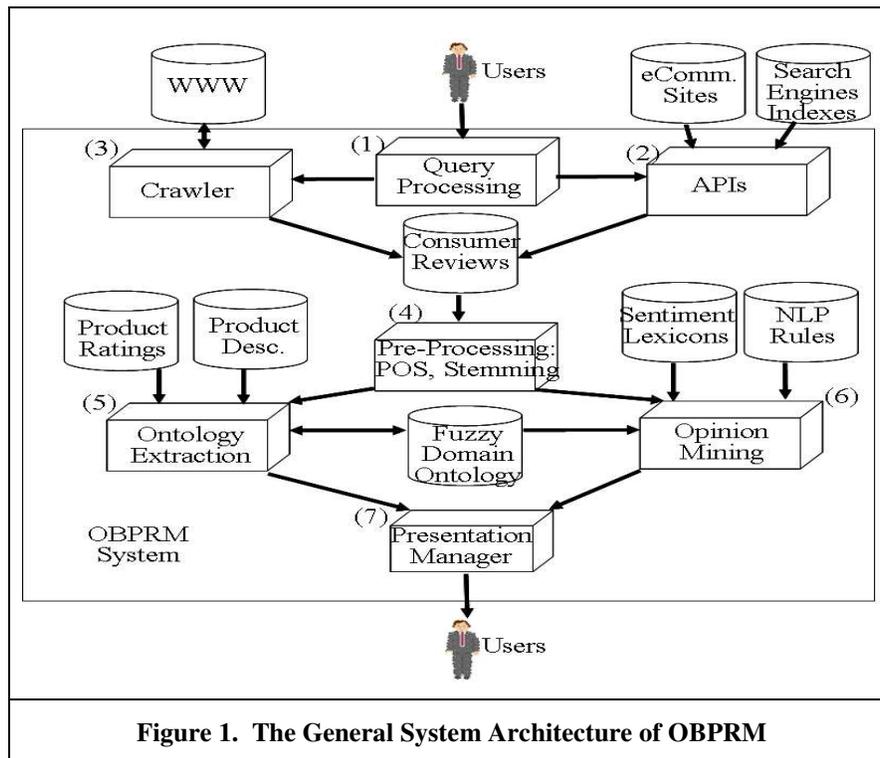
Figure 1. The General System Architecture of OBPRM

Ontology extraction (Task 5 in Figure 1) is carried out offline and it must be performed before context-sensitive mining (Task 6 in Figure 1) is conducted. The fuzzy domain ontology captures taxonomic information such as "iPhone" (product) "is-a" mobile phone (product category), and non-taxonomic relationship such as "screen" (product feature) is "associated with" "iPhone" (product). In addition, context-sensitive sentiment orientation (e.g., "excellent") of a product feature (e.g., "screen") is also captured in the fuzzy domain ontology. Consumer reviews, product ratings, and product descriptions can be retrieved from e-Commerce sites; this information is fed to the ontology extraction module to automatically build the fuzzy domain ontology. The details about ontology extraction will be described in the following section. Based on the fuzzy domain ontology, manually crafted sentiment lexicons, and basic NLP rules, the opinion mining module can analyze each pair of product feature and sentiment (*f,*

---

s) and determine its polarity. By aggregating the polarity scores of product features from all the reviews, a final sentiment score and a polarity label can be generated for the target product. The presentation manager is responsible for delivering the opinion mining results as well as visualizing the fuzzy domain ontology (Task 7 in Figure 1). Our prototype system[6] was developed using Java (J2SE v 1.4.2), Java Server Pages (JSP) 2.1, and Servlet 2.5. The system is hosted on a DELL 1950 III Server with 16GB main memory and running under Apache Tomcat 6.0.

The design choices of OBPRM can be explained based on the merits of proven technologies. Firstly, an ontology based knowledge representation for the automatically generated sentiment lexicon is justified because formal ontology such as the W3C's Web Ontology Language (OWL)[7] facilitates knowledge exchange between humans and systems on the Web (Fikes et. al. 2004). Therefore, representing our sentiment lexicon as domain ontology facilitates the extraction and reuse of the sentiment knowledge across various Web applications. Secondly, as the problem of sentiment lexicon construction is viewed as the process of ontology learning, existing ontology extraction techniques (Lau et. al. 2009; Tho et. al. 2006) can be applied to build a sentiment lexicon automatically. In particular, the notions of fuzzy set and fuzzy relation can be applied to capture the uncertainty presented in the problem domain (Zadeh 1965). Although machine learning techniques have been explored for context-sensitive opinion mining, a large number of manually labeled training examples at the phrase level are often required to train an accurate classifier. Given the sheer volume of user contributed opinion data in the era of Web 2.0, our proposed statistical learning approach which does not rely on manually annotated training examples is desirable; our method can scale up to process the ever growing opinion data on the Web. Finally, our proposed method also utilizes proven IR techniques such as Term Frequency Inverse Document Frequency (TFIDF) (Salton 1990) and Rocchio learning (Rocchio 1971) for product feature extraction, and the Keyword Classifier (Kindo et. al. 1997; Lau et. al. 2008) for sentiment polarity prediction. These methods have been empirically tested and they are efficient enough to process opinionated documents of a Web scale.

Fuzzy Domain Ontology Extraction

Ontology is generally considered as a formal specification of conceptualization which consists of concepts and their relationships (Gruber 1993). Domain ontology is one kind of ontology which is used to represent the knowledge for a particular type of application domain (e.g., a consumer product domain) (Dittenbach et. al. 2004). Our model of fuzzy domain ontology is underpinned by fuzzy sets and fuzzy relations (Zadeh 1965). The fuzzy domain ontology offers the expressive power such that the uncertainty related to sentiment polarity prediction can be properly captured. In particular, the light weight fuzzy domain ontology model is developed based on the formal model published in (Lau et. al. 2009). Our light weight fuzzy domain is defined as follows:

Definition 1. *Fuzzy Set:* A fuzzy set F consists of a set of objects drawn from a domain X and the membership of each object x_i in F is defined by a membership function $\mu_F : X \mapsto [0,1]$.

Definition 2. *Fuzzy Relation:* A fuzzy relation R_{XY} is defined as the fuzzy set R on a domain $X \times Y$ where X and Y are two crisp sets. The membership of each object (x_i, y_i) in R is defined by a membership function $\mu_R : X \times Y \mapsto [0,1]$.

Definition 3. *Fuzzy Domain Ontology:* A fuzzy domain ontology is a triple $Ont = \langle C, R_{NTAX} R_{TAX} \rangle$, where C is a set of concepts (classes). The fuzzy relation $R_{NTAX} : C \times C \mapsto [0,1]$ defines the strength of the non-taxonomic relationship for each pair (c_i, c_i) in R_{NTAX}, and the fuzzy relation $R_{TAX} : C \times C \mapsto [0,1]$ defines the strength of the taxonomic (subclass/super-class) relationship for each pair (c_i, c_i).

With reference to our application, C represents the set of products, product categories, sentiments, and so on. For our application, the taxonomy relations R_{TAX} are specified by the users. When a user inquires about the sentiment of a product, they will select the product category pertaining to the product via our system interface. The main focus of our fuzzy domain ontology extraction method is to automatically learn the non-taxonomic fuzzy relation

---

[6] http://quantum.is.cityu.edu.hk/OM_web/login.jsp

[7] http://www.w3.org/TR/owl-features/

R_{NTAX} (e.g., "Associated") among the classes $c_i \in C$. A conceptual view of the fuzzy domain ontology when it is applied to our opinion mining problem is depicted in Figure 2. Our fuzzy domain ontology can be formally represented by a standard representation language such as OWL.

Figure 2. The Fuzzy Domain Ontology
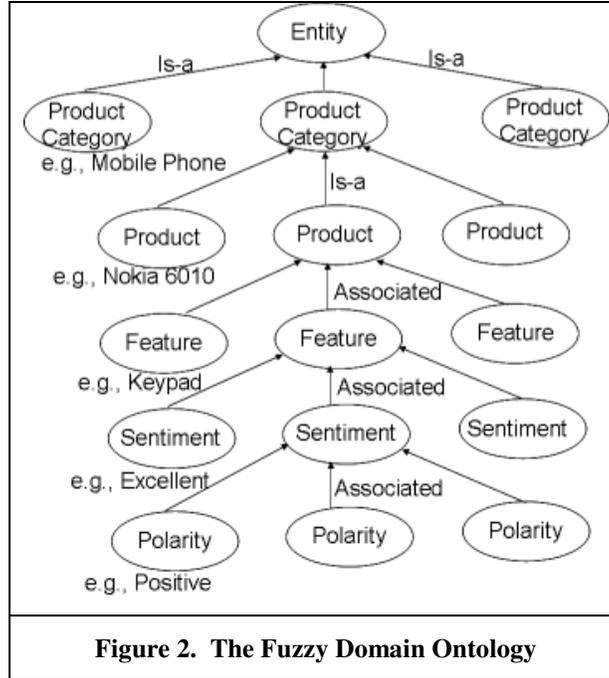
Extracting Product Features

For each product category defined in our system, the set of product features associated with the category is acquired via an offline ontology building process. Since each product belongs to a product category, the common product features associated with each product will implicitly be associated with a product category via the "is-a" relation. Using the APIs provided by e-Commerce sites, our system can retrieve the product descriptions of a set of products under a product category. Crawler programs and the APIs of Google can also be used to collect product descriptions for each relevant product. Standard document-processing procedures (Task 4 in Figure 1) are applied to each product description document retrieved from the Web. Normalized TFIDF weighting scheme (Salton 1990) is applied to extract the most informative noun patterns to represent the product features of a particular product p_i.

For each product description document d, the weight $w(f_i, d) \in [0,1]$ of a product feature f_i is derived by:

$$w(f_i, d) = \frac{\left(0.5 + 0.5 \frac{tf(f_i)}{Maxtf(d)}\right) \cdot \log_2 \frac{N}{df(f_i)}}{\sqrt{\sum_{f_j \in d}\left((0.5 + 0.5 \frac{tf(f_j)}{Maxtf(d)}) \cdot \log_2 \frac{N}{df(f_j)}\right)^2}} \tag{1}$$

where the term $tf(f_i)$ is the term frequency of f_i in d, and $df(f_i)$ is the document frequency of f_i in the collection of product descriptions retrieved from the Web (i.e., how many times f_i occurs in the product descriptions). The function $Maxtf(d)$ returns the maximal term frequency from a product description d. $N = |D_{p_i}|$ is cardinality of the set of product descriptions D_{p_i} retrieved for the product p_i. Finally, the fuzzy membership of a product feature for a product is approximated by the mean TFIDF weights of f_i over the collection of product descriptions D_{p_i}:

$$\mu_{R_{NTAX}}(f_i, p_i) \approx \frac{\sum_{d \in D_{p_i}} w(f_i, d)}{|D_{p_i}|} \qquad (2)$$

During opinion mining of consumer reviews (Task 6 in Figure 1), the product features f_i of each review d are also extracted using the aforementioned process. Let $\vec{d}$ be a vector of product feature weights extracted from a review d. After the opinion mining process, the vectors of product feature weights derived from the set of consumer reviews D_{Rev} are applied to update the $\mu_{R_{NTAX}}(f_i, p_i)$ of the fuzzy domain ontology using an approach similar to the Rocchio learning method (Rocchio 1971):

$$\vec{F}_{i,t+1} = \alpha \times \vec{F}_{i,t} + \frac{\beta}{|D_{Rev}|} \times \sum_{d \in D_{Rev}} \frac{\vec{d}}{\|\vec{d}\|} \qquad (3)$$

where $\vec{F}_{i,t} = < \mu_{R_{NTAX}}(f_1, p_i), \mu_{R_{NTAX}}(f_2, p_i), \cdots, \mu_{R_{NTAX}}(f_n, p_i) >$ is the original vector of product feature weights (i.e., $\mu_{R_{NTAX}}(f_i, p_i)$) for product p_i. The parameters $\alpha = \beta = 0.5$ were applied to our experiments; $\|\vec{d}\|$ is the norm (length) of a product feature vector $\vec{d}$. After Rocchio learning, an updated set of product features and their weights $\vec{F}_{i,t+1}$ is obtained for the product. While new product features may be added, the weakest product features are removed after Rocchio learning. The parameter $\varpi_f = 50$ controls how many product features retained for the product feature vector $\vec{F}_{i,t+1}$.

Extracting Sentiments Related to Product Features

Similar to product feature extraction, a set of consumer reviews is used to build the non-taxonomic relations between sentiments and product features via an offline learning process. The Adjectives or Adverbs associated with the product features (measured by a text window of size ϖ_{win}) within a review are extracted as the candidate sentiments (Subrahmanian and Reforgiato 2008). The meaning of $\varpi_{win} = 1$ is that the adjective or adverb next to product feature from both sides are extracted. Our system takes into account the sentence boundary as well. Within a text window, the negation of sentiment will be taken into account. For instance, if words such as "no", "not", "except", and so on is found, the negation of the sentiment word is assumed (Das and Chen 2007; Ding and Liu 2007). By means of Balanced Mutual Information (BMI) which has successfully been applied to build fuzzy domain ontology (Lau et. al. 2009), we can identify the sentiments that are highly associated with a given product feature:

$$\begin{aligned} \mu_{R_{NTAX}}(s_i, f_i) \quad \approx \quad & BMI(t_i, t_j) \\ = \quad & \varpi_{BMI} \times [Pr(t_i, t_j) \log_2(\frac{Pr(t_i, t_j)+1}{Pr(t_i)Pr(t_j)}) + \\ & Pr(\neg t_i, \neg t_j) \log_2(\frac{Pr(\neg t_i, \neg t_j)+1}{Pr(\neg t_i)Pr(\neg t_j)})] - \\ & (1 - \varpi_{BMI}) \times [Pr(t_i, \neg t_j) \log_2(\frac{Pr(t_i, \neg t_j)+1}{Pr(t_i)Pr(\neg t_j)}) + \\ & Pr(\neg t_i, t_j) \log_2(\frac{Pr(\neg t_i, t_j)+1}{Pr(\neg t_i)Pr(t_j)})] \end{aligned} \qquad (4)$$

where $\mu_{R_{NTAX}}(s_i, f_i)$ is the membership function to estimate the degree of association between a sentiment s_i and a product feature f_i. The advantage of the BMI measure is that it takes into account both term presence and term absence as the evidence of the implicit term association. The parameter $\varpi_{BMI} \in [0.5, 0.7]$ was used to adjust the relative weight of positive and negative evidence respectively (Lau et. al. 2009; Lau 2003). $Pr(t_i, t_j)$ is the joint probability that both terms appear in a text window, and $Pr(t_i)$ is the probability that a term t_i appears in a text

window. The probability $Pr(t_i)$ is estimated based on $\frac{|w_t|}{|w|}$ where $|w_t|$ is the number of windows containing the term t and $|w|$ is the total number of windows constructed from a corpus. Similarly, $Pr(t_i, t_j)$ is the fraction of the number of windows containing both terms out of the total number of windows. After computing the BMI scores, the top $\varpi_s = 30$ sentiments associated with a product feature can be extracted. All the BMI scores are subject to a linear normalization process (i.e., $BMI_{normal} = \frac{BMI - BMI_{Min}}{BMI_{Max} - BMI_{Min}|}$) such that $\mu_{R_{NTAX}}(s_i, f_i) \in [0,1]$ is maintained. The degree of association between the pair (s_i, f_i) is incrementally updated based on $\mu_{R_{NTAX}}^{t+1}(s_i, f_i) = \alpha \times \mu_{R_{NTAX}}^t(s_i, f_i) + \beta \times BMI_{normal}(s_i, f)$ after scanning a new collection of consumer reviews.

Learning Context-Sensitive Sentiment Polarity

Our objective is to develop a fuzzy membership function to estimate $\mu_{R_{NTAX}}((s_i, f_i), o_i)$ given a product p_i and a feature f_i, where $o_i \in \{positive, negative, neutral\}$ is a sentiment orientation label. Instead of relying on manually tagged examples to train a classification function as in (Wilson et. al. 2005), we would like to develop an automated method such that context-sensitive sentiment polarity can be acquired across different product domains. The basic intuition is that a positive consumer review is more likely to contain positive sentiment and feature pairs (s_i, f_i) than a negative review does. Therefore, we may use the sentiment polarity label of a consumer review to infer the sentiment polarity of an individual product feature within the review. Based on the theory of Kullback-Leibler (KL) divergence (Kullback and Leibler 1951), an effective measure called Keyword Classifier (KC) has been developed to identify positive, negative, and neutral keywords representing an information seeker's positive, negative, or neutral information needs (Kindo et. al. 1997; Lau et. al. 2008). Instead of summing the probabilities characterizing the positive and negative events as in the original KL divergence formulation, the KC measure takes a subtraction between the conditional probabilities related to the positive and the negative events. Such a formulation corresponds to our intuition of weighting positive and negative sentiments presented in consumer reviews. For the OBPRM system, we apply the KC measure to determine the polarity and the corresponding strength of a $t = (s_i, f_i)$ pair extracted from a review. As the ratings of consumer reviews are readily available from e-Commerce sites, we can obtain the polarity label for a review by using the respective APIs. For example, an Amazon rating of 4-5 can be regarded as positive, and a rating of 1 can be taken as negative; a mid range rating of 2-3 is considered neutral. The KC formulation is shown as follows (Kindo et. al. 1997; Lau et. al. 2008):

$$KC(t) = \tanh(\frac{df(t)}{\varpi_{pos}} \times \Pr(pos \mid t) \times \log_2 \frac{\Pr(pos \mid t)}{\Pr(pos)} - \frac{df(t)}{\varpi_{neg}} \times \Pr(neg \mid t) \times \log_2 \frac{\Pr(neg \mid t)}{\Pr(neg)}) \tag{5}$$

$$polarity_{Ont}(t) = \begin{cases} \dfrac{KC(t) - \varpi_{KC}}{1 - \varpi_{KC}} & \text{if } KC(t) > \varpi_{KC} \\[2ex] -\left(\dfrac{|KC(t)| - \varpi_{KC}}{1 - \varpi_{KC}}\right) & \text{if } KC(t) < -\varpi_{KC} \\[2ex] 0 & \text{otherwise} \end{cases} \tag{6}$$

The parameters ϖ_{pos} and ϖ_{neg} control the learning rates for positive and negative evidences respectively, and they can be established empirically (Kindo et. al. 1997; Lau et. al. 2008). The hyperbolic tangent function tanh ensures the induced polarity scores fall in the unit interval. $\Pr(pos|t) = \frac{df(t_{pos})}{df(t)}$ is the estimated conditional probability that a review is positive given that it contains the particular sentiment feature pair $t = (s_i, f_i)$; it can be derived based on a set of consumer reviews (i.e., the context). $\Pr(pos|t)$ is estimated based on the fraction of the number of positive

reviews which contain the pair (i.e., $df(t_{pos})$) over the total number of reviews which contain the pair (i.e., $df(t)$). Similarly, $\Pr(neg|t) = \dfrac{df(t_{neg})}{df(t)}$ is the estimated conditional probability that a review is negative if it contains the pair *t*. The document frequency $df(t_{neg})$ represents the number of negative reviews which contain the pair *t*. In addition, $\Pr(pos) = \dfrac{|D^{+}_{Rev}|}{|D^{+}_{Rev}| + |D^{-}_{Rev}|}$ ($\Pr(neg) = \dfrac{|D^{-}_{Rev}|}{|D^{+}_{Rev}| + |D^{-}_{Rev}|}$) is the priori probability that a review is positive (negative) respectively; D^{+}_{Rev} (D^{-}_{Rev}) is the set of positive (negative) reviews rated by consumers. A positive $polarity_{Ont}(t)$ score indicates that the underlying pair of sentiment and product feature is positive, whereas a negative $polarity_{Ont}(t)$ score implies that the pair is negative. If the polarity score is zero, the pair is considered neutral. Similarly, the membership function can be incrementally updated by $\mu^{t+1}_{R_{NTAX}}((s_i, f_i), o_i) = \alpha \times \mu^{t}_{R_{NTAX}}((s_i, f_i), o_i) + \beta \times polarity_{Ont}(s_i, f)$ at training stage $t+1$.

Context-Sensitive Opinion Mining

Identification of Product Features and Sentiments

Given a fuzzy domain ontology which holds the context-sensitive sentiment polarities of product features, it is possible to improve the accuracy of opinion mining (Task 6 in Figure 1). The main input to the OBPRM system is a user's query about a product. Driven by such a query, all the consumer reviews are retrieved by means of OBPRM's APIs and crawlers. Before sentiment analysis is applied to each consumer review, standard document pre-processing procedures (e.g., stop word removal, POS tagging, and stemming) are applied to the review document (Task 4 in Figure 1). As illustrated in the previous section, normalized TFIDF weighting is applied to extract the most representative product features from each review. In addition, low frequency candidate product features (Hu and Liu 2004; Yang et. al. 2007) are identified by matching the target tokens with the common product features stored in the fuzzy domain ontology. Candidate sentiments which are close (within a text window of size ϖ_{win}) to the product features are then identified and selected according to the normalized BMI scores. As mentioned in the previous section, the negation of a sentiment word is taken into account by our system if a negation indicator such as "no", "not", "except", and so on is found in the same text window of the sentiment (Das and Chen 2007; Ding and Liu 2007).

Predicting the Polarities of Reviews and Products

Our system first applies the fuzzy domain ontology to determine the strength and the polarity for each (s_i, f_i) pair. If there is a sentiment polarity that cannot be resolved, the system will apply the linguistic rules (Ding and Liu 2007) stored in the NLP rule base to determine the sentiment polarity. If the NLP rules cannot be applied to the current target, a default sentiment lexicon such as SentiWordNet (Esuli and Sebastiani 2005) is invoked to determine the pair's polarity. If no classification can be done after utilizing all the sentiment sources, a neutral polarity will be assigned to the (s_i, f_i) pair. The overall polarity score $polarity_{D_{Rev}}(s_i, f_i)$ of the pair (s_i, f_i) over a collection of consumer reviews D_{Rev} is the mean polarity score computed based on the weighted polarity scores generated from individual reviews. The polarity score $polarity_{doc}(d)$ for a review *d* is derived by:

$$polarity_{doc}(d) = \frac{\sum\limits_{(s_i, f_i) \in d} \varpi_{source} \times polarity(s_i, f_i)}{|d|} \qquad (7)$$

where $polarity(s_i, f_i) ::= polarity_{Ont}(s_i, f_i) | polarity_{NLP}(s_i, f_i) | polarity_{Lexicon}(s_i)$ represents the context-sensitive sensitive polarity score derived from the fuzzy domain ontology $polarity_{Ont}(s_i, f_i)$, or the polarity score inferred based on our linguistic rules and the seeding sentiment words $polarity_{NLP}(s_i, f_i)$, or the context-free polarity score determined based on the generic sentiment lexicons $polarity_{Lexicon}(s_i)$. If the polarity of a pair (s_i, f_i) is defined in

the system generated sentiment lexicon, the polarity inferred based on the linguistic rules module or the manual lexicon module will be ignored. The weight factor ϖ_{source} defines the importance of the polarity score when it is extracted from different sources e.g., system generated lexicon with weight ϖ_{Ont}, linguistic rules with weight ϖ_{NLP}, or generic sentiment lexicon with weight $\varpi_{Lexicon}$. If a negation indicator is associated with the pair (s_i, f_i), the sign of the polarity score $polarity(s_i, f_i)$ will be reversed. The parameters $(\varpi_{Ont} = \varpi_{NLP} = 1, \varpi_{Lexicon} = 0.5)$ are used to tune the relative weights of the polarity scores generated from the respective sources. The term $|d|$ returns the cardinality of d in terms of the number of sentiment and product feature pair found in d. Three simple linguistic rules proposed by (Ding and Liu 2007) and fourteen seeding sentiment words used by (Turney and Littman 2003) are used to build our NLP module for polarity detection. Finally, the polarity score $polarity_{Pro}(p_i)$ for a product p_i is derived by:

$$polarity_{Pro}(p_i) = \frac{\sum\limits_{f_i \in F(p_i), d \in D_{Rev}, (s_i, f_i) \in d} \sum \mu_{R_{NTAX}}(f_i, p_i) \times \mu_{R_{NTAX}}(s_i, f_i) \times polarity(s_i, f_i)}{|D_{Rev}| \times |d|} \tag{8}$$

where $\mu_{R_{NTAX}}(f_i, p_i)$ is the product and product feature association weight (i.e., the fuzzy membership) defined in the fuzzy domain ontology and $\mu_{R_{NTAX}}(s_i, f_i)$ is the fuzzy membership of the relation (s_i, f_i). The polarity score $polarity(s_i, f_i)$ of a (s_i, f_i) pair is computed based on the fuzzy domain ontology, the NLP module, or the sentiment lexicon. A threshold $\varpi_{Pro} = 0.05$ is established empirically to determine the ultimate polarity of the product. If the product score $polarity_{Pro}(p_i)$ is greater than ϖ_{Pro}, the product will be labeled as positive; if the product score is leas than $-\varpi_{Pro}$, the product will be labeled as negative; otherwise it is considered neutral.

Experiments and Results

Similar to previous studies (Archak et. al. 2007; Hu and Liu 2004), we retrieved real consumer reviews from Amazon.com using the Amazon Web services APIs. Our evaluation work was based on eight Amazon product categories such as Cameras, Mobile Phones, Watches, Laptops, Sport Equipment, and so on. The average length of these reviews is 139.4 words, and the average number of unique words per product category is 34,549.4 words. For each product category, 5,000 consumer reviews together with the corresponding product ratings (in the scale of 1 to 5) were downloaded. So, our dataset included 40K reviews in total. For our experiments, we treated the ratings of 4-5 as positive and the rating of 1 as negative. The evaluation metrics included precision, recall, accuracy, and F-measure (van Rijsbergen 1979):

$$precision = \frac{a}{a+b} \tag{9}$$

$$recall = \frac{a}{a+c} \tag{10}$$

$$F_{\beta=1} = \frac{(1+\beta^2) \times precision \times recall}{\beta^2 \times precision + recall} = \frac{2a}{2a+b+c} \tag{11}$$

$$accuracy = \frac{a+d}{a+b+c+d} \tag{12}$$

With reference to a confusion matrix, *a, b, c, d* refer to the number of correctly classified positive (negative) reviews, the number of classified non-positive (negative) reviews, the number of non-classified positive (negative) reviews, and the number of non-classified non-positive (negative) reviews. Ten-fold cross validation was employed to evaluate our system (Mitchell 1997). In each experimental run, one randomly selected sub-sample (a fold) was retained as the validation set and the remaining 9 sub-samples were used as the training set (Archak et. al. 2007).

This procedure was repeated 10 times and we took the average result as the respective measurement score. With our novel automated domain ontology learning method, manual tagging of the consumer reviews is not required.

Experiment One

In the first experiment, we tried to compare the performance between context-sensitive and context-free opinion mining methods. For each product pertaining to some consumer reviews, we used our crawler to retrieve the product feature or the product description section from the respective product page of Amazon.com. (Eq. 1) and (Eq. 2) were then used to select at most fifty features (i.e., ϖ_f =50) with each feature represented by a single noun. (Eq. 4-6) were applied to learn the sentiment orientations from the training set of reviews; the automatically acquired contextual sentiment knowledge is stored in our fuzzy domain ontology. Based on the empirical testing for several sub-samples, the parameters ($\varpi_{pos} = 100, \varpi_{neg} = 10, \varpi_{KC} = 0.02$) were established. After learning the fuzzy domain ontology in each run, we applied (Eq. 7) to predict the polarities of the reviews of the validation set. To facilitate the comparison with published classification results (Hu and Liu 2004; Popescu and Etzioni 2005), we also adopted a two-class (i.e., positive and negative) classification procedure. If the polarity score $polarity_{doc}(d)$ of a review d was greater than zero, the review was classified as positive; otherwise it was negative. For the baseline runs (context-free opinion mining), ontology learning was not invoked. Polarity prediction was conducted based on the NLP module and the sentiment lexicon module only.

Table 1. Comparative Performance Between the Baseline System and OBPRM					
Product Category	Baseline System		OBPRM		Improvement	
	F-Measure	Accuracy	F-Measure	Accuracy	F-Measure	Accuracy
Toys	0.7672	0.6588	0.8345	0.7562	+8.78%	+14.78%
Sport Equipment	0.7683	0.6616	0.8274	0.7526	+7.70%	+13.75%
Watches	0.8091	0.7044	0.8871	0.8260	+9.64%	+17.26%
Laptops	0.7994	0.6922	0.8918	0.8342	+11.55%	+20.51%
Motorcycle Parts	0.7845	0.6920	0.8624	0.8044	+9.93%	+16.24%
Cameras	0.7928	0.6872	0.8740	0.8100	+10.25%	+17.87%
MP3 Players	0.7830	0.6902	0.8718	0.8174	+11.33%	+18.43%
Mobile Phones	0.7630	0.6616	0.8479	0.7856	+11.12%	+18.74%
Average	0.7834	0.6810	0.8621	0.7983	**+10.04%**	**+17.22%**

Table 1 tabulates the F-measure and the accuracy results pertaining to various product categories for the baseline system and the ontology-based opinion mining system (OBPRM) respectively. Figure 3 plots the comparative performance between the two systems in terms of accuracy scores across the eight product categories. As shown in Table 1, the average improvements across the eight product categories are +10.04% and +17.22% in terms of F-measure and Accuracy respectively. The performance (F-measure and Accuracy) of the OBPRM system is significantly better than that of the baseline system according to Wilcoxon signed rank test (*p*<0.01). The product category of laptop computers has the largest improvement in terms of Accuracy and F-measure. The intuitive explanation of this improvement is that sentiments such as "small", "tiny", "little", and so on usually have negative polarities defined in sentiment lexicons such as SentiWordNet (Esuli and Sebastiani 2005) or OpinionFinder (Wilson et. al. 2005). However, for the context of laptop computers, these sentiments often imply a positive orientation (e.g., "a small laptop consuming little time for configuration"). According to our data analysis, 60.5% of the sentiments found in the laptop category are "weak subjective" as defined according to OpinionFinder. These weak subjective sentiments (e.g., "tiny") are usually context-dependent, and so they become the source of performance improvement after our context-sensitive sentiment polarity detection.
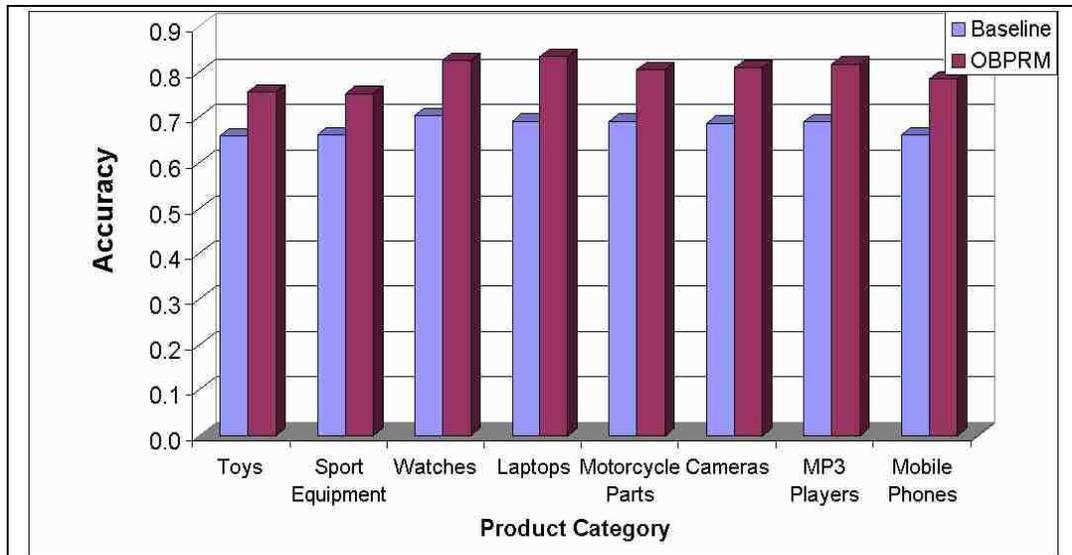
Figure 3. Comparative Performance Between OBPRM and the Baseline

Table 2. Context-Sensitive Sentiments for the Product Category of Laptops					
Positive Sentiments			**Negative Sentiments**		
Features	Sentiments	Polarity	Features	Sentiments	Polarity
drive	hard	0.999997	battery	defective	-0.85361
gb	hard	0.999567	screen	blue	-0.83926
keyboard	great	0.997659	laptop	hopeless	-0.82427
laptop	great	0.996959	problem	called	-0.82290
laptop	small	0.996951	product	average	-0.79484
battery	long	0.995394	light	lit	-0.78244
battery	great	0.992897	warranty	damaged	-0.78244
computer	great	0.989026	laptop	blue	-0.73699
work	great	0.986654	hp	called	-0.73208
size	small	0.969728	acer	average	-0.73208
screen	size	0.965520	support	called	-0.73208
battery	good	0.964125	laptop	damaged	-0.71784
drive	external	0.963565	product	defective	-0.67862
keyboard	easy	0.959373	customer	called	-0.67216
keyboard	good	0.959208	machine	frustrating	-0.63849
computer	small	0.952062	no	stopped	-0.61123
machine	great	0.949557	problem	searching	-0.61031
life	great	0.946326	product	faulty	-0.60185
weight	light	0.945798	wireless	hopeless	-0.60185
ram	hard	0.942497	pixels	poor	-0.60185

Samples of the product features, sentiments, and the corresponding polarity scores captured in our fuzzy domain ontology are illustrated in Table 2. Only the top 20 positive and negative sentiments are depicted in the table. As shown in Table 2, it is interesting to find that the usual negative sentiment "small" as defined in generic sentiment lexicons (Esuli and Sebastiani 2005; Wilson et. al. 2005) is shown to be positive (e.g., small size laptop) in the contexts of Laptops. Another interesting finding is that general product name such as "laptop" which is not usually regarded as product feature is treated as an implicit feature from the perspective of the review writers. Therefore,

when a product feature based sentiment lexicon is built, explicit or implicit product features extracted from consumer reviewers should be used to substitute the initial product features extracted from generic product description by applying an appropriate learning mechanism such as (Eq.3).

Experiment Two

Earlier studies indicated that representation schemes of product features and the proximity consideration for sentiments identification might have impact on the accuracy of opinion mining (Hu and Liu 2004; Popescu and Etzioni 2005). Our second experiment examined the impact of product feature representation (with varying length of noun patterns) and sentiment identification (with different proximity to the identified product feature) on the overall performance of context-sensitive opinion mining. We tried a proximity factor from 1 (sentiments immediately next to the identified product feature on both sides) to 10 words, and representing product features by "Noun (N)", "Noun Noun (NN)", or "Noun Noun Noun (NNN)" patterns. The other parameters and the dataset were the same as those used in experiment one. Figure 4 depicts the average accuracy of the OBPRM system with varying proximity factor and length of product feature for the 8 product categories. It is shown that a unigram (single noun) representation of product feature and a proximity factor of 5 to 6 words produce the best opinion mining accuracy. Consistent with the previous observation (Hu and Liu 2004), a proximity factor of 5 words is effective for extracting sentiments related to the identified product features. With a small proximity factor, many candidate sentiments were missed. On the other hand, a large proximity factor might introduce too many irrelevant tokens (noises) to the process of sentiment extraction. However, unlike the previous studies (Hu and Liu 2004; Dave et. al. 2003; Popescu and Etzioni 2005), we found that unigram (a single noun) was more effective than bigram (two nouns) or trigram (three nouns) for the representation of product features. According to our in-depth analysis, it was revealed that standard names of product features such as "battery life" for mobile phones or laptops were not frequently used in consumer reviews. Indeed, in more than a half of the reviews related to mobile phones, the word "battery" alone (e.g., "the battery can last for 5 days") was referred to. As a result, quite a number of important product features were not extracted from the reviews if we used the "Noun Noun" pattern to represent product features. Similarly, as most of the names of product features comprise one to two words, using the pattern of "Noun Noun Noun" to represent product features will lead to a very poor recall. In fact, a previous study also found that a single noun is sufficient to represent a product feature (Archak et. al. 2007). Irrespective of the word length of the product features, it seems that a proximity factor of five to six words is appropriate according to our empirical study.

Figure 4. The Impact of Feature Representation and Proximity Factor on Prediction Accuracy
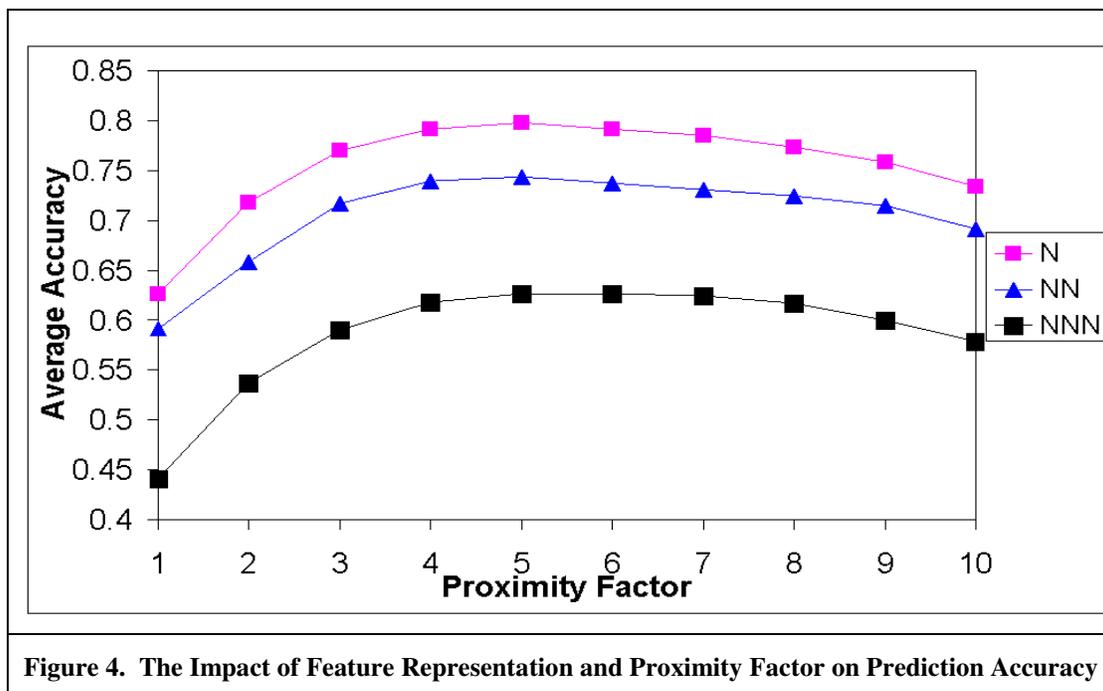
Experiment Three

For our last experiment, we tried to evaluate the performance of the OBPRM system based on the publicly available benchmark dataset originally retrieved from Amazon.com (Hu and Liu 2004). The benchmark dataset[8] consists of the consumer reviews for five products such as cameras, mobile phones, and MP3 players. Each sentence of a review was manually tagged by two researchers, but the actual consumer rating of the review was not included. For instance, the tagged sentence "camera[+2]##this is a great camera for you!" means that the feature "camera" is positive with a polarity score 2. As shown in this example, a feature tagged in the benchmark dataset may not be an explicit product feature at all. In order to evaluate our system based on this benchmark dataset, we treated each sentence which had a polarity tag as a review. Ten-fold cross validation procedure was still applied in this experiment. If the polarity (e.g., positive or negative) classified by our system is the same as the manually tagged polarity in the benchmark dataset, it is considered a correct classification. The performance of our system and that of the published results are tabulated in Table 3. It is shown that the precision, recall, and F-measure scores of our system outperform that of FBS (Hu and Liu 2004) and OPINE (Popescu and Etzioni 2005) and it is comparable with that of Opinion Observer (Ding and Liu 2007). Our system achieves the top F-measure score because of the automatic learning of proper sentiment contextual knowledge during the training phrase. The contextual knowledge was then applied to bootstrap opinion mining during the testing phrase. It should be noted that our system does not rely on the manually tagged polarity label to bootstrap classification performance; it only requires the readily available consumer ratings at e-Commerce sites to infer the polarities of sentiments presented in consumer reviews.

Table 3. Comparative Performance Based on Benchmark Test			
Systems	Recall	Precision	F-Measure
FBS	0.693	0.642	0.667
OPINE	0.890	0.860	0.875
Opinion Observer	0.910	0.920	0.910
OBPRM	0.913	0.919	0.916

According to published results (Abbasi et. al. 2008; Pang et. al. 2002), the accuracy of Support Vector Machine (SVM) based sentiment polarity classification falls in the range of [0.8, 0.9]. Although a direct comparison between our work and early studies is not possible due to varying experimental settings, the accuracy of our context-sensitive opinion mining method (e.g., 0.83 for the laptop category and 0.89 for the benchmark dataset) is comparable with that of the early work which utilizes machine learning techniques.

Conclusions and Future Work

Guided by the design science research methodology, we illustrate the design, development, and evaluation of a novel fuzzy domain ontology based context-sensitive opinion mining system in this paper. In particular, we show that a domain specific sentiment lexicon can be automatically constructed to facilitate context-sensitive opinion mining based on existing fuzzy domain ontology extraction method. By applying a variant of the Kullback-Leibler divergence statistical learning technique, our system can accurately predict the polarities of sentiments without requiring extra human effort to annotate training examples. Based on real consumer reviews collected from Amazon.com, the effectiveness of our OBPRM context-sensitive opinion mining system is empirically tested; the proposed system performs significantly better than a baseline system which is based on context-free sentiment classification approach. Experimental results from a benchmark test also reveal that the performance of the OBPRM system is better than that of similar opinion mining systems. The business implication of our research is tremendous; our context-sensitive opinion mining methodology assists organizations to analyze a large number of consumer reviews efficiently. As a result, organizations can develop effective business strategies related to marketing, customer support, and product design functions in a timely fashion. Our work also facilitates individual consumers' comparison shopping processes. Future research involves examining the correlation of the product sentiment scores generated by our system and the actual sales ranks or sales volumes of various products. Another direction of research is to examine the validity (e.g., review spam) and the usefulness of reviews posted to the Web. A direct comparison of the performance of our method with that of some machine learning classification methods such as SVM will be conducted in the future.

---

[8] http://www.cs.uic.edu/~liub/FBS/CustomerReviewData.zip

Acknowledgements

References

Abbasi, A., Chen, H., and Salem, A. "Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums", *ACM Transactions on Information Systems* (26:3), June 2008, Article 12.

Agarwal, A., Biadsy, F., and Mckeown, F. "Contextual Phrase-Level Polarity Analysis Using Lexical Affect Scoring and Syntactic N-Grams", in *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics,* 2009, pp.24-32.

Archak, N., Ghose, A., and Ipeirotis, P. "Show me the money!: deriving the pricing power of product features by mining consumer reviews", in *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Jose, California, 12-15 August 2007, pp. 56 - 65.

Bao, S., Li, R., Yu, Y., and Cao, Y. "Competitor Mining with the Web", *IEEE Transactions on Knowledge and Data Engineering,* (20:10), 2008, pp. 1297–1310.

Bickart, B. and Schindler, R. "Internet forums as influential sources of consumer information", *Journal of Interactive Marketing*, (15:3), 2001, pp. 31–40.

Cimiano, P., Hotho, A., and Staab, S. "Learning concept hierarchies from text corpora using formal concept analysis", *Journal of Artificial Intelligence Research* (24), 2005, pp. 305–339.

Danescu-Niculescu-Mizil, C., Kossinets, G., Kleinberg, J., and Lee, L. "How opinions are received by online communities: a case study on amazon.com helpfulness votes", in *Proceedings of the 18th international conference on World Wide Web*, Madrid, Spain, 20-24 April 2009, pp.141-150.

Das, S. and Chen, M. "Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web", *Management Science*, (53:9), September 2007, pp. 1375–1388.

Dave, K., Lawrence, S., and Pennock, D. "Mining the peanut gallery: opinion extraction and semantic classification of product reviews", in *Proceedings of the 12th international conference on World Wide Web*, Budapest, Hungary, 20-24 May 2003, pp.519-528.

Ding, X. and Liu, B. "The Utility of Linguistic Rules in Opinion Mining", in *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Amsterdam, Netherlands, 23-27 July 2007, pp. 811 - 812.

Dittenbach, D., Berger, H., and Merkl, D. "Improving domain ontologies by mining semantics from text", in *Proceedings of the First Asia-Pacific Conference on Conceptual Modelling (APCCM2004)*, 2004, pp. 91–100.

Esuli, A. and Sebastiani, F. "Determining the semantic orientation of terms through gloss classification", in *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, Bremen, Germany, October 2005, pp. 617 - 624.

Fikes, R., Hayes, P., and Horrocks, I. "OWL-QL - A Language for Deductive Query Answering on the Semantic Web", Journal of Web Semantics, (2:1), 2004, pp. 19-29.

Gruber, T.R. "A translation approach to portable ontology specifications", *Knowledge Acquisition*, (5:2), 1993, pp. 199–220.

Hatzivassiloglou, V. and McKeown, K. "Predicting the semantic orientation of adjectives", in *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, 1997, pp. 174 - 181.

Hevner, A., March, S., Park, J., Ram, S. "Design science in information systems research", *MIS Quarterly* (28:1), 2004, pp. 75-105.

Hu, M. and Liu, B. "Mining and summarizing customer reviews", in *Proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Seattle, Washington, 22-25 August 2004, pp. 168-177.

Kindo, T., Yoshida, H., Morimoto, T., and Watanabe, T. "Adaptive personal information filtering system that organises personal profiles automatically", in *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence (IJCAI'97)*, 1997, pp. 716-721.

Kullback, S. and Leibler, R.A. "On Information and Sufficiency", *The Annals of Mathematical Statistics*, (22:1), 1951, pp. 79–86.

Lau, R.Y.K. "Context-Sensitive Text Mining and Belief Revision for Intelligent Information Retrieval on the Web",

Journal of Web Intelligence and Agent Systems, (1:3-4), 2003, pp. 151-172.

Lau, R.Y.K., Bruza, P.D., and Song, D. "Towards a Belief Revision Based Adaptive and Context Sensitive Information Retrieval System", *ACM Transactions on Information Systems*, (26:2), March 2008, pp. 8:1-8:38.

Lau, R.Y.K., Song, D., Li, Y., Cheung, C.H., Hao, J.X. "Towards A Fuzzy Domain Ontology Extraction Method for Adaptive e-Learning", *IEEE Transactions on Knowledge and Data Engineering*, (21:6), 2009, pp. 800-813.

Lau, R.Y.K. and Lai, C.L. "Information Granulation for the Design of Granular Information Retrieval Systems", in *Proceedings of the 2008 International Conference on Information Systems*, Paris, France, 14-17 December 2008, Paper 179.

Liu, Y., Huang, X., An, A., and Yu, X. "ARSA: A Sentiment- Aware Model for Predicting Sales Performance Using Blogs", in *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Amsterdam, Netherlands, 23-27 July 2007, pp. 607-614.

Lu, Y. and Zhai, C. "Opinion integration through semi-supervised topic modeling", in *Proceedings of the 17the International Conference on World Wide Web*, Beijing, China, 21-25 April 2008, pp. 121-130.

Macdonald, C. and Ounis, I. "Overview of the TREC 2007 Blog Track", in *Proceedings of the Sixteenth Text REtrieval Conference*, Gaithersburg, Maryland, 2007. Available from http://trec.nist.gov/pubs/trec16/.

Miao, Q., Li, Q., and Dai, R. "An integration strategy for mining product features and opinions", in *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, Napa Valley, California, 26-30 October 2008, pp. 1369-1370.

Miller, G., Beckwith, R., Fellbaum, C., Gross, D., and Miller, K. "Introduction to WordNet: An on-line lexical database", *Journal of Lexicography*, (3:4), 1990, pp. 234–244.

Mitchell, T. Machine Learning. New York: McGraw-Hill, 1997.

Pang, B., Lee, L., and Vaithyanathan, S. "Thumbs up? Sentiment Classification using Machine Learning Techniques", in *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing*, Philadelphia, 2002, pp. 79-86.

Porter, M. "An algorithm for suffix stripping", *Program*, 1980, (14:3), pp.130–137.

Popescu, A.M. and Etzioni, O. "Extracting Product Features and Opinions from Reviews", in *Proceedings of the 2005 Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, Vancouver, Canada, October 2005, pp. 339-346.

Rocchio, J. "Relevance Feedback in Information Retrieval", in G. Salton, editor, *The SMART retrieval system: experiments in automatic document processing*, 1971, pp. 313-323.

Salton, G. and McGill, H.J. Introduction to Modern Information Retrieval. New York: McGraw-Hill, 1983.

Salton, G., Wong, A., and Yang, C.S. "A vector space model for automatic indexing", *Communications of the ACM* (18:11), November 1975, pp. 613-620.

Salton, G. "Full text information processing using the smart system", *Database Engineering Bulletin*, (13:1), March 1990, pp. 2–9.

Stone, P., Dunphy, D., Smith, M., and Ogilvie, D. The General Inquirer: A Computer Approach to Content Analysis. Cambridge: MIT Press, 1966.

Subrahmanian, V. and Reforgiato, D. "AVA: Adjective-Verb-Adverb Combinations for Sentiment Analysis", *IEEE Intelligent Systems*, July/August 2008, pp. 43-50.

Tan, S., Wang, Y., and Cheng, X. "Combining learn-based and lexicon-based techniques for sentiment detection without using labeled examples", in *Proceedings of the 31st annual international ACM SIGIR Annual International Conference on Research and Development in Information Retrieval*, Singapore, 20-24 July 2008, pp.743-744.

Tan, S., Wu, G., Tang, H., and Cheng, X. "A novel scheme for domain-transfer problem in the context of sentiment analysis", in Proceedings of the sixteenth ACM conference on Conference on information and knowledge management, Lisboa, Portugal, 6-8 November 2007, pp.979-982.

Tho, Q.T., Hui, S.C., Fong, A., and Cao, T.H. "Automatic Fuzzy Ontology Generation for Semantic Web", *IEEE Transactions on Knowledge and Data Engineering*, (18:6), June 2006, pp. 842-856.

Turney, P. and Littman, M. "Measuring praise and criticism: Inference of semantic orientation from association", *ACM Transactions on Information Systems*, (21:4), October 2003, pp. 315–346.

van Rijsbergen, C. Information Retrieval. London:Butterworths, 1979.

Whitelaw, C., Garg, N. and Argamon, S. "Using appraisal groups for sentiment analysis", in *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, Bremen, Germany, October 2005, pp. 625 - 631.

Wille, R. "Formal concept analysis as mathematical theory of concepts and concept hierarchies", in Bernhard Ganter, Gerd Stumme, and Rudolf Wille, editors, *Formal Concept Analysis, Foundations and Applications*,

volume 3626, 2005, pp. 1–33.

Wilson, T., Wiebe, J., and Hwa, R. "Just how mad are you? finding strong and weak opinion clauses", in *Proceedings of the 19th National Conference on Artificial Intelligence*, San Jose, California, 25-29 July 2004, pp. 761–769.

Wilson, T., Wiebe, J., and Hoffmann, P. "Recognizing contextual polarity in phrase-level sentiment analysis", in *Proceedings of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, 2005, pp. 347–354.

Wilson, T., Wiebe, J., and Hwa, R. "Recognizing strong and weak opinion clauses", *Computational Intelligence* (22:2), 2006, pp. 73-99.

Wright, A. "Our sentiments, exactly", *Communications of the ACM*, (52:4), April 2009, pp. 14-15.

Xu, R., Wong, K.F., Lu, Q., Xia, Y., and Li, W. "Learning Knowledge from Relevant Webpage for Opinion Analysis", in *Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, Sydney, Australia, 9-12 December 2008, pp. 307–313.

Yang, K., Yu, N., and Zhang, H. "WIDIT in TREC-2007 Blog Track: Combining lexicon based Methods to Detect Opinionated Blogs", in Proceedings of the Sixteenth Text REtrieval Conference, Gaithersburg, Maryland, 2007, Available from http://trec.nist.gov/pubs/trec16/.

Zadeh, L.A.. "Fuzzy sets", *Journal of Information and Control*, (8), 1965, pp. 338-353.

Zhang, M. and Ye, X. "A generation model to unify topic relevance and lexicon-based sentiment for opinion retrieval", in *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Singapore, 20-24 July 2008, pp. 411-418.