Opinion Extraction and Summarization on the Web

Minqing Hu and Bing Liu

Department of Computer Science
University of Illinois at Chicago
851 South Morgan Street,
Chicago, IL 60607-7053
{mhu1, liub}@cs.uic.edu

Abstract

The Web has become an excellent source for gathering consumer opinions. There are now numerous Web sources containing such opinions, e.g., product reviews, forums, discussion groups, and blogs. Techniques are now being developed to exploit these sources to help organizations and individuals to gain such important information easily and quickly. In this paper, we first discuss several aspects of the problem in the AI context, and then present some results of our existing work published in KDD-04 and WWW-05.

Introduction

The Web has dramatically changed the way that people express their opinions. They can now express their views on almost anything in review sites, Internet forums, discussion groups, and blogs. This online word-of-mouth behavior represents new and valuable sources of information for marketing intelligence and many other applications. Techniques are now being developed to exploit these sources to help companies, organizations and individuals to gain such information effectively and easily.

In this paper, we use consumer reviews of products as an example to explore this important problem, which is not only academically challenging but also very useful in practice (Hu and Liu 2004; Liu, Hu and Cheng 2005).

Merchants selling products on the Web often ask their customers to review the products that they have purchased and the associated services. There are also many dedicated review sites. With more and more people writing reviews, the number of reviews grows rapidly. There are also a large number of sites containing reviews, which make it hard for users to read and to analyze them.

Given a set of customer reviews of a particular product, we propose to study the following problems.

1. Identifying product features that customers have expressed their opinions on,
2. For each feature, identifying review sentences that give positive or negative opinions, and
3. Producing a feature-based summary of opinions on the product using the discovered information.

Let us use an example to illustrate the task. For example, we have the review sentence "*The picture quality of this camera is amazing*". The product feature that has been commented on is "*picture quality*". The opinion expressed in the sentence by the reviewer is positive. Based on such information from a large number of reviews, we can produce the *feature-based summary* in Figure 1.

*Digital_camera*_1:
Feature: **picture quality**
Positive: 253
<individual review sentences>
Negative: 6
<individual review sentences>
Feature: **size**
Positive: 134
<individual review sentences>
Negative: 10
<individual review sentences>
…

Figure 1: An example summary

In Figure 1, picture quality and (camera) size are the product features. The number after positive/negative is the number of positive/negative opinions expressed on the feature. The <individual review sentences> link points to the specific sentences and/or the whole reviews that give positive/negative comments about the feature.

Using such summaries of different products, we can compare consumer opinions on competing products. Figure 2 shows a comparison visually, which compares customer opinions on two digital cameras along different feature

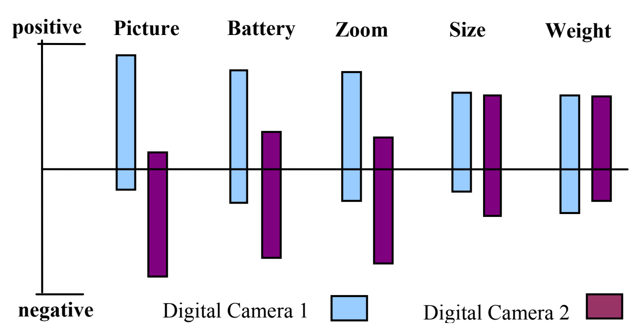

Figure 2: Visual comparison of opinions on two products

dimensions, i.e., picture, battery, zoom, etc. Each bar in Figure 2 shows the percent of reviews that express positive (above x-axis) or negative (below x-axis) opinions on a feature. With a single glance of the visualization, the user can clearly see the strengths and weaknesses of each product in the minds of consumers.

In this paper, we discuss this problem in the context of several areas of AI research and describe some of our existing approaches and results.

Problem in the AI Context

The problem is basically a natural language processing or a text mining problem. There are several related AI research areas. We discuss them below.

Sentiment Classification

Sentiment classification classifies a review or an evaluative text as positive or negative. Several researchers have worked on the problem. Works of Sack (1994) uses models inspired by cognitive linguistics. Das and Chen (2001) use a manually crafted lexicon and several scoring methods to classify stock postings. Turney (2002) applies an unsupervised learning technique based on the mutual information between document phrases and the words “excellent” and “poor”, where the mutual information is computed using statistics gathered from a search engine. Pang *et al*. (2002) examine several machine learning methods for sentiment classification of movie reviews. A good evaluation of various methods for sentiment classification is also given in (Dave, Lawrence and Pennock 2003) based on classification of reviews. Using available training corpus from some Web sites, the authors design and experiment a number of methods for building sentiment classifiers. They also show that such classifiers performed quite well with test reviews, but not on sentences. Morinaga *et al.* (2002) study the problem of finding the reputation of the target product. (Yi *et al* 2003) studies a similar problem. Other related work included (e.g., Wilson, Wiebe and Hwa 2004; Riloff, Wiebe and Phillips 2005; Kim and Hovy 2004).

Much of the existing work focuses on whole review classification. Opinion extraction discussed in this paper goes much further. It studies reviews at the sentence and product feature level, which we believe are more useful in practice. Specifically, opinion extraction differs from most existing works in two main aspects: (1) the new focus is not on classifying each review as a whole but on classifying each sentence in a review that contains some product feature. Within a review some sentences may express positive opinions about certain product features while some other sentences may express negative opinions about some other product features. (2) Little existing research has been done on identifying product features that have been commented on by reviewers. This information is very useful in practice. We presented several initial approaches to solve the problems in (Hu and Liu 2004; Liu, Hu and Cheng 2005). More recent work on the topic includes (e,g., Popescu and Etzioni 2005; Carenini, Ng and Zwart 2005).

Information Extraction from Text

Our task of identifying product features from reviews is also related to information extraction, or more specifically, terminology finding because product features are usually terminologies. In terminology finding, there are two main techniques: symbolic approaches that rely on syntactic description of terms, namely noun phrases, and statistical approaches that exploit the fact that the words composing a term tend to be found close to each other and reoccurring (Jacquemin and Bourigault 2001; Justeson and Katz 1995; Daille 1996). However, our experiments show that using noun phrases tends to produce too many non-terms (low precision), while using reoccurring phrases misses many low frequency terms, terms with variations, and terms with only one word. We experimented with the term extraction and indexing system, FASTR[1]. However, the results were poor. Both the recall and precision of FASTR were significantly lower than those of our method.

For general information extraction, (Mooney and Bunescu 2005) gives a good survey of current techniques. The main extraction problem that has been studied extensively is the named entity extraction in various contexts. (Mooney and Bunescu, 2005) reports that Conditional Random Fields performed the best. We also tried Conditional Random Fields for extracting product features from a specific type of reviews, in which Pros and Cons are short phrases or incomplete sentences (Liu, Hu and Cheng 2005). However, the results were not satisfactory. We believe that the main reason is that named entities usually have strong indicators, e.g., capitalization of words. Product features have few strong indicators.

Synonym Identification

One issue of product feature extraction is that it generates a large number of features and many of them are synonyms. For example, “picture” and “image” in a camera review have the same meaning and thus should be associated together. In order to produce the feature-based summary described in the Introduction section, it is crucial to identify and to group synonyms together. The main existing methods of finding synonyms or lexical similarity of words can be categorized into four categories: using a monolingual dictionary, using WordNet (Fellbaum 2001), using a thesaurus, and computing the mutual information of a pair of words by using co-occurrence counts obtained via queries to a search engine (Turney 2001). We have examined the method of using WordNet to find synonyms. However, the result is very limited. More research is this area is needed. We believe that context analysis will play a major role as most synonyms of product features are context or product dependant.

[1] http://www.limsi.fr/Individu/jacquemi/FASTR/

Text Summarization

Our feature-based summarization is related to classic text summarization. However, opinion summary is a structured summary instead of a short paragraph of free text.

Most existing text summarization works are based on passage extraction (Paice 1990; Kupiec, Sparck-Jones 1993; Hovy and Lin 1997), i.e., extracting some important sentences and using them as the summary. In the past, the focus was on a single document summary. In recent years, researchers also studied summarization of multiple documents covering similar information. Their main purpose was to summarize the similarities and differences of the information content in these documents (Mani and Bloedorn 1997; Radev, McKweown 1998). Opinion summarization is quite different as its aim is to find the product features that have been commented on in multiple reviews. The summary is produced based on the positive and negative review counts on the product features.

Our Work

In this section, we briefly discuss our existing work. We begin by describing different types of reviews on the Web. We then present some initial approaches and results.

There are three main review formats on the Web. Different review formats may need different techniques to perform the opinion extraction task.

Format (1) - Pros and Cons: The reviewer is asked to describe Pros and Cons separately.

Format (2) - Pros, Cons and detailed review: The reviewer is asked to describe Pros and Cons separately and also write a detailed review.

Format (3) - free format: The reviewer can write freely, i.e., no separation of Pros and Cons.

For formats (1) and (2), opinion orientations (positive or negative) of features are known because Pros and Cons are separated. Thus only product features need to be identified. For format (3), we need to identify both product features and opinion orientations. In both formats (1) and (3), reviewers typically use full sentences. However, for format (2), Pros and Cons tend to be very brief. For example, under Cons, one may write: "heavy, bad picture quality", which are elaborated in the detailed review.

We used both data mining and natural language processing techniques to extract product features. We have studied all three formats. Formats (1) and (3) are basically the same except that there is no need to determine opinion orientations (positive or negative) for format (1). We now give a brief overview of our existing techniques.

Extracting Product Features

We used different techniques for extracting features from free format reviews (format (1) and format (3)) and from brief Pros and Cons (format (2)).

In both cases, we first run NLProcessor[2] to generate part-of-speech (POS) tags. We then used data mining techniques to extract product features.

[2] NLProcessor – http://www.infogistics.com/textanalysis.html

Free format reviews (formats (2) and (3)): After POS tagging is done, we first identify frequent nouns or noun phrases as features using association rule mining. The frequent features are then used to find potential opinion adjective words, which are employed to extract associated infrequent features (Hu and Liu, 2004).

Pros and Cons (format (2)): The above feature extraction method does not work well for Pros and Cons, as it cannot find implicit features (features that are not nouns). Thus, we proposed a supervised mining method to generate language patterns from training reviews and then use these patterns to extract product features from test reviews. In (Liu, Hu and Cheng 2005), class association rule mining is used (Liu, Hsu and Ma, 1998), which can produce patterns such as "<NN> [feature] <VB> is <JJ> *any_word*." This pattern is employed to extract the noun as a product feature if it is followed by verb "is" and an adjective. See (Liu, Hu and Cheng 2005) for more details.

Using the generated language patterns to extract features from short phrases gives quite accurate results. However, they do not do so well on complete sentences in reviews of formats (1) and (3). The main reason is that the mining system generates too many patterns. It is hard to know which pattern to use given a sentence.

Opinion Orientation Prediction

For reviews of format (3), we need to identify opinion sentences in each review and decide whether each opinion sentence is positive or negative. Note that these opinion sentences must contain one or more product features identified above. To decide the opinion orientation of each sentence, we perform three sub-tasks. First, a set of opinion words (adjectives, as they are normally used to express opinions) is identified. If an adjective appears near a product feature in a sentence, then it is regarded as an opinion word. Second, for each opinion word, we determine its semantic orientation, i.e., positive or negative. We use WordNet (Fullbaum 2001) in conjunction with a set of seed words to determine whether an opinion word is positive or negative. Finally, we decide the opinion orientation of each sentence by using its dominant orientation. We are still studying other classification and NLP methods for the task.

Evaluations

We collected and annotated two corpora from the Web for two types of reviews to evaluate our techniques.

Corpora

Corpus1 consists of customer reviews of five electronics products from Amazon.com: 2 digital cameras, 1 DVD player, 1 mp3 player, and 1 cellular phone. These reviews are of format (3).

Corpus2 contains a collection of reviews of 15 electronic products from Epinions.com in format (2). 10 of them are used in training to mine language patterns. The patterns are then applied to extract features from test reviews of the rest 5 products (Pros and Cons are considered separately).

Results

Corpus1: For this corpus, we need to perform both feature extraction and opinion orientation identification. For details, please refer to (Hu and Liu 2004)

For product feature extraction, we obtained an average recall and precision of 0.80 and 0.72 respectively. We note that these results are only based on features being nouns and noun phrases. In practice, verbs and adjectives can also indicate product features, which were not included in this work and in the experiment results. We considered them in the experiment with Corpus2.

For opinion orientation identification, we evaluate our system from two perspectives:

1. Effectiveness of opinion sentence identification. We obtained the recall and precision of 0.69 and 0.64.
2. Effectiveness of opinion orientation prediction of opinion sentences. We obtained the accuracy of 0.84.

Corpus2: For this corpus, we only need to extract product feature as their opinion orientations are already known due to the separation of Pros and Cons. The average recall and precision for Pros are 0.90 and 0.89. For Cons, the average recall and precision are 0.82 and 0.79. More experimental result can be found in (Liu, Hu and Cheng 2005).

Conclusions

In this paper, we discussed the problem of opinion extraction and summarization in the context of product reviews. We believe that the same tasks can also be performed on postings in Internet forums and discussion groups. We have shown that the problem touches several AI research areas and is both intellectually challenging and practical useful. We also briefly described some of our existing approaches and their results based on two corpora that we have collected on the Web. In our future work, we will further improve and refine our techniques, and to deal with the outstanding problems, i.e., pronoun resolution, determining the strength of opinions, and investigating opinions expressed with adverbs, verbs and nouns. In the past year, several other researchers also started to work on the problem (e,g., Popescu and Etzioni 2005; Carenini, Ng and Zwart 2005). We believe that with more researchers studying this problem, it is possible to build industrial strength systems in the near future.

References

Alm, C., Roth, D. and Sproat, R. Emotions from text: machine learning for text-based emotion prediction. *HLT/EMNLP 2005*.

Carenini, G. Ng, R. and Zwart, E. 2005. Extracting knowledge from evaluative text. *K-CAP 05*.

Daille. 1996. Study and Implementation of Combined Techniques for Automatic Extraction of Terminology. *The Balancing Act: Combining Symbolic and Statistical Approaches to Language Processing*. MIT Press, Cambridge

Das, S. and Chen, M., 2001. Yahoo! for Amazon: Extracting market sentiment from stock message boards. *APFA'01*.

Dave, K., Lawrence, S., and Pennock, D., 2003. Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews. *WWW-03*.

Hearst, M, 1992. Direction-based Text Interpretation as an Information Access Refinement. In Paul Jacobs, editor, *Text-Based Intelligent Systems*. Lawrence Erlbaum Associates.

Fellbaum, C. (Edit). WordNet, An Electronic Lexical Database. 2001. The MIT Press

Hovy, E., and Lin, C.Y. 1997. Automated Text Summarization in SUMMARIST. *ACL Workshop on Text Summarization*.

Hu, M. and Liu, B. 2004. Mining and summarizing customer reviews. *KDD'04*.

Jacquemin, C., and Bourigault, D. 2001. Term extraction and automatic indexing. In R. Mitkov, editor, *Handbook of Computational Linguistics*. Oxford University Press.

Justeson, J. S., and Katz, S.M. 1995. Technical Terminology: some linguistic properties and an algorithm for identification in text. *Natural Language Engineering* 1(1):9-27.

Kim, S and Hovy, E. Determining the sentiment of opinions. *COLING'04*, 2004.

Kupiec, J., Pedersen, J., and Chen, F. 1995. A Trainable Document Summarizer. *SIGIR-95*.

Liu, B., Hsu, W., and Ma, Y. 1998. Integrating Classification and Association Rule Mining. *KDD-98*, 1998.

Liu, B., Hu, M., and Cheng, J. Opinion Observer: Analyzing and Comparing Opinions on the Web. *WWW-05*.

Mani, I., and Bloedorn, E., 1997. Multi-document Summarization by Graph Search and Matching. *AAAI-97*.

Mooney, R. J. and Bunescu, R. Mining Knowledge from Text Using Information Extraction. *SIGKDD Explorations* 7, 2005.

Morinaga, S., Ya Yamanishi, K., Tateishi, K, and Fukushima, T. 2002. Mining Product Reputations on the Web. *KDD'02*.

Paice, C. 1990. Constructing Literature Abstracts by Computer: techniques and prospects. *Info. Proc. & Mang,* 26:171-186.

Pang, B., Lee, L., and Vaithyanathan, S., Thumbs up? Sentiment Classification using machine learning techniques. *EMNLP-02*.

Popescu, A. and Etzioni, O. Extracting product features and opinions from Reviews. *EMNLP-05*

Radev, D. and McKeown, K. 1998. Generating natural language summaries from multiple on-line sources. *Computational Linguistics*, 24(3):469-500, September 1998.

Riloff, E. Wiebe, J. Phillips, W. Exploiting Subjectivity Classification to Improve Information Extraction, *AAAI*-2005.

Sack, W., 1994. On the Computation of Point of View. *AAAI'94*.

Sparck-Jones, K. 1993b. What might be in a summary? *Information Retrieval* 93: 9-26.

Turney, P. 2001. Mining the Web for synonyms: PMI-IR versus LSA on TOEFL, *ECML-01*.

Turney, P. Thumbs up or thumbs down? semantic orientation Applied to Unsupervised Classification of Reviews. *ACL* 2002.

Wilson, T, Wiebe, J, & Hwa, R. Just how mad are you? Finding strong and weak opinion clauses. *AAAI*-2004.

Yi, J., Nasukawa, T., Bunescu, R., Niblack, W. Sentiment Analyzer: Extracting sentiments about a given topic using natural language processing techniques. ICDM-2003.