

OMS-J: An Opinion Mining System for Japanese Weblog Reviews Using a Combination of Supervised and Unsupervised Approaches

Guangwei Wang

Graduate School of Information
Science and Technology
Hokkaido University
Sapporo, Japan 060-0814
wgw@media.eng.hokudai.ac.jp

Kenji Araki

Graduate School of Information
Science and Technology
Hokkaido University
Sapporo, Japan 060-0814
araki@media.eng.hokudai.ac.jp

Abstract

We introduce a simple opinion mining system for analyzing Japanese Weblog reviews called OMS-J. OMS-J is designed to provide an intuitive visual GUI of opinion mining graphs for a comparison of different products of the same type to help a user make a quick purchase decision. We first use an opinion mining method using a combination of supervised (a Naive Bayes Classifier) and unsupervised (an improved SO-PMI: Semantic Orientation Using Pointwise Mutual Information) learning.

1 Introduction

Nowadays, there are numerous Web sites containing personal opinions, e.g. customer reviews of products, forums, discussion groups, and blogs. Here, we use the term Weblog for these sites. How to extract and analyze these opinions automatically, i.e. “Opinion Mining”, has seen increasing attention in recent years.

This paper presents a simple opinion mining system (OMS-J) for analyzing Japanese Weblog reviews automatically. The novelty of OMS-J is two-fold: First, it provides a GUI using intuitive visual mining graphs aimed at inexperienced users who want to check opinions on the Weblog before purchasing something. These graphs can help the user to make a quick decision on which product is suitable. Secondly, this system combines a supervised and an unsupervised approach to perform opinion mining. In related work (Chaovalit, 2005; Turney, 2002), both supervised and unsupervised approaches have been shown to have their pros and

cons. Based on the merits of these approaches and the characteristics of Japanese (Kobayashi, 2003), we proposed an opinion mining method using a Naive Bayes Classifier (supervised approach) and an improved SO-PMI method (unsupervised approach) to perform different parts of the classification task (Wang, 2006).

OMS-J implements Weblog opinion mining by the steps shown in Figure 1. In the next section, we describe the proposed system in detail.

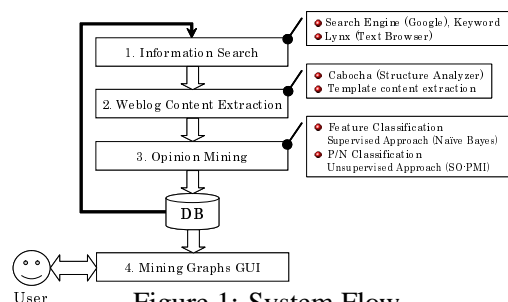


Figure 1: System Flow

2 Proposed System

2.1 Information Search

The first step is information search. We used the Google search engine¹ to get all the information on one product category or one specific product in the Japanese weblog on the Internet. The search keyword is the product category name or the product name. The URL range of the search is restricted by the URL type (e.g. blog, bbs, review).

2.2 Weblog Content Extraction

The Content Extraction step first analyzes the Weblog content using a dependency structure analyzer for Japanese, Cabocha². Based on the syntactic characteristics of Japanese reviews and the results

¹<http://www.google.co.jp/>

²<http://www.chasen.org/~taku/software/cabocha/>

of related work (Kobayashi, 2003; Taku, 2002), we designed the following templates to extract opinion phrases:

- < noun + auxiliary word + adj / verb / noun >
- < adj + noun / undefined / verb >
- < noun + verb >
- < noun + symbol + adj / verb / noun >
- Except the above < adj >

2.3 Opinion Mining

Opinion mining methods can usually be divided into two types: supervised and unsupervised approaches. Supervised approaches are likely to provide more accurate classification results but need a training corpus. Unsupervised approaches on the other hand require no training data but tend to produce weaker results.

We propose a combined opinion mining method by performing feature classification and P/N classification (Wang, 2006). The purpose of these classifications is to know what the opinion expresses about a certain product's features. Feature means a product's attribute, i.e. price, design, function or battery feature. Based on our previous study, it is easy to create a feature corpus. Therefore feature classification is performed by a supervised approach, a Naive Bayes Classifier. P/N classification classifies reputation expressions into positive or negative meaning using an unsupervised approach, SO-PMI. The SO-PMI approach measures the similarity of pairs of words or phrases based on the mutual information theory, in our case the closeness of an opinion and words for "good" or "bad".

No human effort is required when mining a new product or category. Only inputting the name of the product or category is needed. It does however require quite a lot of processing time, since the SO-PMI approach using a search engine is very time consuming. Adding new features requires manual work, since a small hand labeled training corpus is used. Similar categories of products, for instance cameras and mp3 players, use the same features though, so this is not done very often.

2.4 Mining Graphs GUI

Finally, OMS-J provides a GUI with mining graphs showing the opinion mining data in the database, as shown in Figure 2. These graphs show the distribution of positive and negative opinions of each feature

type such as "design", and for each product. The distribution of positive opinions among the different product choices are shown in a pie chart, as is the same for negative opinions. This GUI can also show graphs for a single product's mining results, showing the positive/negative opinion distribution of each feature.

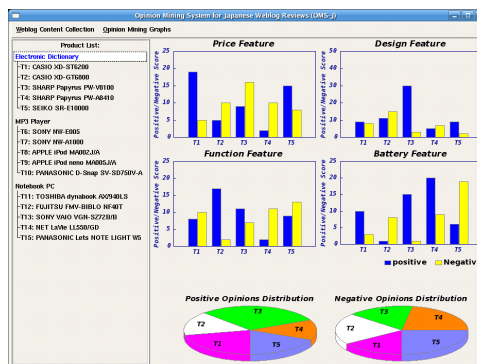


Figure 2: OMS-J's GUI Screenshot for One Product Category

3 Demonstration

During the demonstration, we will show that OMS-J is an intuitive opinion mining system that can help people to make a quick decision on purchasing some product. OMS-J's trial version has been developed and tested with three kinds of products: Electronic Dictionaries, MP3 Players and Notebook PCs. The experiment results were positive. We will show how the system works when a user wants to buy a good MP3 player or wants to get a feel for the general opinions on a specific Notebook PC etc.

References

- Pimwadee Chaovalit and Lina Zhou. 2005. *Movie Review Mining: a Comparison between Supervised and Unsupervised Classification Approaches*. Proceedings of the 38th Annual HICSS.
- Peter D. Turney. 2002. *Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews*. Proceedings 40th Annual Meeting of the ACL, pp. 417-424.
- Nozomi Kobayashi, Kentaro Inui, Yuji Matsumoto, Kenji Tateishi and Toshikazu Fukushima. 2003. *Collecting evaluative expressions by a text mining technique*. IPSJ SIG NOTE, Vol.154, No.12.
- Guangwei Wang and Kenji Araki. 2006. *A Decision Support System Using Text Mining Technology*. IEICE SIG Notes W12-2006-6, pp. 55-56.
- Taku Kudoh and Yuji Matsumoto. 2002. *Applying Cascaded Chunking to Japanese Dependency Structure Analysis*. Information Processing Society of Japan (IPSJ) Academic Journals, Vol 43, No 6, pp. 1834-1842.