A Graphical User Interface for Feature-Based Opinion Mining

Pedro Balage Filho
University of Wolverhampton
pedrobalage@gmail.com

Caroline Brun
Xerox Research Centre Europe
Caroline.Brun@xrce.xerox.com

Gilbert Rondeau
Xerox Research Centre Europe
Gilbert.Rondeau@xrce.xerox.com

Abstract

In this paper, we present XOpin, a graphical user interface that have been developed to provide a smart access to the results of a feature-based opinion detection system, build on top of a parser.

1 Introduction

Opinion mining (or sentiment analysis) arouses great interest in recent years both in academia and industry. Very broadly, sentiment analysis aims to detect the attitude of a person toward a specific topic expressed in natural language and to evaluate the polarity of what is been expressed, i.e., whether it is positive or negative. With the emergence of the Web 2.0, i.e., forums, blogs, web sites compiling consumer reviews on various subjects, there is a huge amount of documents containing information expressing opinions: the "user generated content". This constitutes a very important data source for monitoring various applications (business intelligence, product and service benchmarking, technology watch). Numerous research works at the crossroads of NLP and data mining are focusing on the problem of opinion detection and mining. In this paper, we present the advanced research prototype we have designed: it consists in an integration of a feature-based opinion detection system together with a graphical user interface providing to the end-user a smart access to the results of the opinion detection.

We first present an overview of sentiment analysis. Then, we detail the system we have developed, in particular the graphical user interface, and conclude.

2 Analyzing Sentiment in Texts

Sentiment Analysis plays a very important role to help people to find better products or to compare product characteristics. For the consumer, a good interface allows to navigate, compare and identify the main characteristics of the products or companies. For the company, it is interesting to know the customer preferences. It is an essential step to optimize marketing campaigns and to develop new features in products.

Despite the increase of interest in sentiment analysis, many tools do not pay much attention to the user interface aspects. These aspects are very important in order to satisfy the user needs.

In the literature, we find some different ways to aggregate and represent the summary information from a collection of texts annotated with sentiment. For instance, Gamon et al. (2005) use colors to display the general assessment of product features. The system shows the reviews as boxes, where the box size indicates the number of mentions of that topic and the color indicates the average sentiment it contains. This interface allows having a quick glance about the most important topics and the sentiment expressed.

Another display idea is presented in the Opinion Observer (Liu et al., 2005). In this system, a bar shows the polarity related with each product and each feature. The portions of the bar above and below a horizontal line represent the amount of positive and negative reviews. For example, in a cell phone domain, the sentiment associated with features like LCD, battery, reception and speaker are used to compare the relevance of one product in opposite to another.

Morinaga et al. (2002) present an interface where the sentiment information is represented by the degrees of association between products and opinion-indicative terms. The author uses principal component analysis to produce a two-dimensional visualization where the terms and products are plotted indicating the relatedness among the points.

In the internet, we can find many systems and companies related with sentiment analysis. For example, the company Lexalytics has in its website

Proceedings of the NAACL-HLT 2012: Demonstration Session, pages 5–8,
Montréal, Canada, June 3-8, 2012.

an available demo[1] for sentiment detection. This demo shows an interface which highlights positive and negative words in the text. The interface also shows entities, categories associated, a summary and the top terms.

The RankSpeed [2] is a website for product comparison. The website includes in the search the sentiment associated with each product. In the interface, the user can input a list of sentiment words, like "excellent", "cool", "easy" or "powerful" that the system will organize the results according the frequency of those words in reviews related to the products.

The Stock Sonar[3] has a timeline chart as the main interface. In this timeline, both positive and negative sentiments are displayed throughout time. The sentiments are retrieved from real-time news associated with a particular company. In the same timeline, it is possible to follow-up the increase or decrease of the stock prices for that company in that period of time. In this application, the sentiment is used to forecast market actions such as buy and sell stocks.

All those systems presented relevant components for a powerful opinion mining interface, but none of them deliver a full interface to explore the multi-aspects in opinion mining. For us, a complete system should provide both single and multi-document visualization, work on the feature level classification, and produce an integrated interface to browse, navigate, filter and visualize files, features and sentiment tendencies. In the following section, we present XOpin, a graphical user interface that have been developed to provide the characteristics described.

3 The System and its Interface

To detect opinions in texts, our system relies on a robust incremental parser, XIP, (Ait-Mokhtar and Chanod 2002), specifically adapted for opinion detection. The system extracts opinions related to the main concepts commented in reviews (e.g. products, movies, books...), but also on features associated to these products (such as certain characteristics of the products, their price, associated services, etc...). More precisely, we adopt the formal representation of an opinion proposed by Liu (2010): an opinion is represented as a five place predicate of the form $(o_j, f_{jk}, so_{ijkl}, h_i, t_l)$, where: o_j is the target of the opinion (the main concept), f_{jk} is a feature associated to the object o_j, so_{ijkl} is the value (positive or negative) of the opinion expressed by the opinion holder h_i about the feature f_{jk}, h_i is the opinion holder, t_l is the time when the opinion is expressed.

We use the robust parser to extract, using syntactic relations already extracted by a general dependency grammar, semantic relations instantiating this model. Other systems use syntactic dependencies to link source and target of the opinion, for example in Kim and Hovy (2006). Our system belongs to this family, as we believe that syntactic processing of complex phenomena (negation, comparison and anaphora) is a necessary step to perform feature-based opinion mining. Another specificity of our system is a two level architecture based on a generic level, applicable to any domain, and on a domain-dependent level, adapted for each sub-domain of application. Regarding evaluation, the relations of opinion extracted by the system have been used to train a SVM classifier in order to assess the system's ability to correctly classify user's reviews as positive or negative. Results are quite satisfying, as they show 93% of accuracy to classify reviews about printers and 89% of accuracy to classify reviews about movies (Brun, 2011).

The XOpin Interface was developed to provide an easy way to allow the user to explore the results of this sentiment analysis system. The interface provides a graphical environment that allows the user to browse, navigate, filter and visualize the necessary information in a collection of texts.

The tool accepts as input pure text files or xml files. The xml files follow a specific format which allows the system to retrieve metadata information. It is also possible to retrieve web pages from the web. The tool offers the possibility to retrieve a single webpage, given the URL, or a collection of pages by crawling. To crawl, for example, reviews webpages, the user need to setup some crawling and information extraction rules defined by a template in the configuration file. The files retrieved from the web are converted in xml format, which allows preserving the metadata information. As an example, Figure 1 shows the

[1]http://www.lexalytics.com/webdemo
[2]http://www.rankspeed.com/
[3]http://www.thestocksonar.com/

organization of this xml file from a review retrieved from the website epinions.com (http://www.epinions.com).

```
<review>
  <source value="http://..." />
  <domain value="Printers"/>
  <brand value="Hewlett Packard"/>
  <product value=" Hewlett Packard 6500A"/>
  <opinion_holder value="user_name"/>
  <review_date value="01/Dec/2011"/>
  <opinion value="Yes"/>
  <review_stars value="5"/>
  <review_popularity value="10"/>
  <textblock layout="title">
        Review Title
  </textblock>
  <textblock layout="summary">
        Review Summary
  </textblock>
  <textblock layout="text">
        Review Free Comment
  </textblock>
</review>
```

Figure 1. Organization of the XML file

The tag *source* keeps the URL from where the review was extracted. The tags domain, brand and product keep the specific data about to the product. The tag *opinion_holder* keeps the name of the user who wrote the review. The tag *review_date* keeps the date when the review was written. The tag *opinion* keeps the user general assessment about the product. In the website epinions.com, the user can assess the product as recommended (Yes) or not recommended (No). The tag *review_stars* contains the number of stars the user attributed to the product. The tag *review_popularity* keeps the number of positive evaluations (thumbsUp) of this particular review by the other users. In the reviews from the website epinions.com we don't have this assessment, so this number represents how many users assigned to trust in this reviewer. The tags *textblock* contain the text for the sections title, summary and review.

After loading a file or a corpus into the tool, the texts are showed in a tree structure in the left panel. A hierarchical structure allows the user to have the corpus organized as a conventional folder structure. In this way, it is possible to analyze the texts inside a specific folder and also to include the texts in the subfolders inside.

To analyze this data, the tool presents three main views: text, timeline and comparison. In the text view, negative terms, positive terms and entities present in the text are highlighted. The purpose of this view is to provide a visual assessment about the sentiment expressed in the text. If the text was loaded by crawling or by an xml file, the metadata is also displayed. Figure 2 shows an example of reviews collected from the website epinions.com, in the category printers.

As said before, XOpin is able to identify the predicates associated with each sentiment and the category it belongs. For example, in the sentence "This printer gives excellent quality color", the tool highlights the positive sentiment "excellent", the predicate associated "color" and organize this predicate into the category color. This predicate categorization depends of the sub-domain architecture level.

This classification is very important to present an organized summary about which category is most positive and with is most negative in the text. The right panel shows this information.

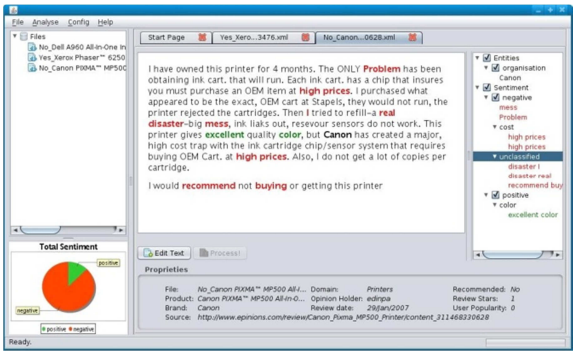

Figure 2. Text visualization in XOpin

The timeline screen (Figure 3) offers the user the option to analyze a corpus of texts organized by time, for example, reviews crawled from the web. In this way, the user can create flexible and interesting views about the products and features found in the corpus.

The timeline shows the total of positive and negative words in the texts for a given date. With this information and a larger enough corpus of reviews it is possible to have a big picture about the user preferences and dissatisfactions.

The timeline also offers the possibility to show the positive and negative lines for specific brands,

products and features in a determined timespan. Filters can remove anything that it is not useful and create a pure visualization about what the user need to see. The left and bottom panels offer options to create those views.

These views can show an evolution in the user's perspective in respect to some new improvement in the product. For example, in a marketing campaign, the company can evaluate the user behavior about the product price.

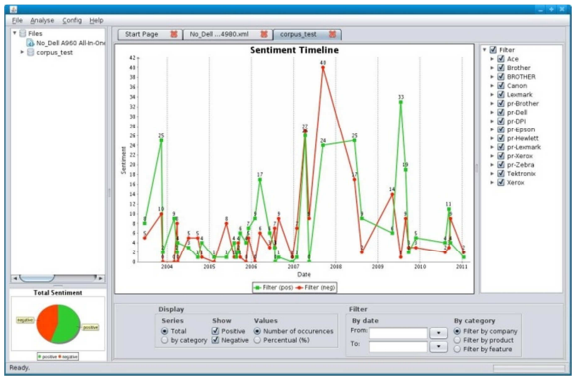

Figure 3. Timeline visualization in XOpin

The comparison view (Figure 4) allows the user to compare side by side different product features in a collection of texts. In this view, the user has the main predicate associated with each feature and the number of positive or negative occurrences. This is interesting in order to have a big picture about what the users are commenting in positive or negative aspects for each feature.

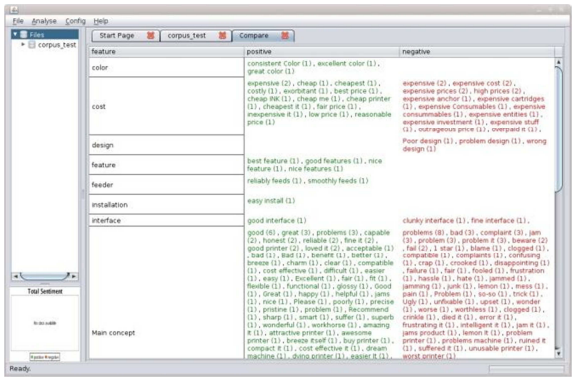

Figure 4. Feature Comparison in XOpin

4 Conclusion

This paper presents an NLP-based opinion mining advanced prototype integrating a dedicated graphical user interface which provides a smart access to the results of the opinion detection. The interface has been build in order to ensure advanced functionalities such as opinion highlighting on text and features, timeline visualization and feature comparison. The system has been demonstrated to potential customers and it received a good feedback. In our assessment, the integrated features provided by the system increased the usability in the data exploration for a reviews corpus compared against other products.

References

Salah Ait-Mokthar, Jean-Pierre Chanod. Robustness beyond Shallowness: Incremental Dependency Parsing. *Special Issue of NLE Journal*, 2002.

Caroline Brun. Detecting Opinions Using Deep Syntactic Analysis. In *Proceedings of the Recent Advances in Natural Language Processing (RANLP)*, Hissar, Bulgaria, September 12-14, 2011.

Michael Gamon, Anthony Aue, Simon Corston-Oliver, and Eric Ringger. Pulse: Mining customer opinions from free text. In *Proceedings of the International Symposium on Intelligent Data Analysis (IDA)*, number 3646 in Lecture Notes in Computer Science, pages 121–132, 2005.

Kim, S.M. and E.H. Hovy. Identifying and Analyzing Judgment Opinions. *Proceedings of the Human Language Technology/HLT-NAACL*. New York, 2006.

Bing Liu, Minqing Hu, and Junsheng Cheng. Opinion observer: Analyzing and comparing opinions on the web. In *Proceedings of WWW*, 2005.

Bing Liu. *Sentiment Analysis and Subjectivity*, Chapter of Handbook of Natural Language Processing, 2nd edition, 2010.

Satoshi Morinaga, Kenji Yamanishi, Kenji Tateishi, and Toshikazu Fukushima.Mining product reputations on the web. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, pages 341–349, 2002.