Scientific Bulletin of the „Petru Maior" University of Tîrgu Mureş
Vol. 9 (XXVI) no. 1, 2012, ISSN 2285 – 438X (Online), ISSN–L 1841 – 9267

THE ORGANIZATION OF AN EFFECTIVE ASPECT-BASED OPPINION MINING SYSTEM

Claudia Maria GOGA[#1], Johann STAN[#2]

[#1] *Department of Mathematics and Informatics,*
Faculty of Sciences and Letters
„Petru Maior" University of Tîrgu Mureş, Romania
Nicolae Iorga Street, no. 1, 540088, Mureş County
[1]goga.claudia@gmail.com

[#2]*Université Jean-Monnet, Saint-Etienne, France*
[2]johann.stan@univ-st-etienne.fr

ABSTRACT

Social platforms are currently used as independent tools where people exchange information about their personal life and share common interests. On the other hand, these people use the Web to browse and find other information, e.g. an interesting destination for holidays, a doctor, a movie, a mobile phone, etc. Even with the emergence of the Web 2.0 - where a user can benefit from other users' experience on the Web, users are still constrained to combine and synthesize themselves information issued from social platforms: read through all of the reviews, check if a user can be trusted or not etc. In this work, we investigate how to build a system that can help reducing the effort needed to synthetize opinions about an item and present it in a user friendly way with new visual metaphors. We focus on the general approach and describe the main steps for building such a framework.

Keywords: opinion mining, sentiment analysis, aspect-based opinion mining, aspect extraction, summarization

1. Introduction

When we need to make a decision like buying a product we often seek opinion from other people. A

blogs or social networks, there are many opinions available on the internet and a product could have hundreds of reviews.

Search engines like Google are suitable for searching for facts, not opinions. Reading them, because there are so many opinions, spread acroass multiple web sites, it is difficult and time consuming.

In order to solve this problem an aggregation and summarization of the opinions of a product is needed.

Opinion mining reffers to classifying an opinion as positive, negative or neutral. An opinion could be expressed about different aspects of the

based-opinion mining offers advantages to both customers and manufacturers:

- Help in making a buying decision and choosing a product over another.

few years ago you would have asked for advise from friends or family. Nowadays, with the growth of social media like online shops,
entity in cause, and for some aspects the sentiment orientation is positve and for others negative. Also a sentence can contain multiple and different opinions on different aspects, knowing that a document or a sentence in a document is positive or negative is not sufficient, thus a more fine-grained level of analysis is needed.

Aspect - based opinion mining represents analysing wheter the opinions expressed about some components or attributes of the product, event or organisation under evaluation are positive or negative.

The tools for aspect

- The time required to analyse opinions of other customers is reduced.
- Help manufacturer receive a feedback on how their products are beeing perceived.

- Help manufacturers improve their products based on the feedback received.
- The user doesn`t need to browse multiple sources of opinions because the reviews are aggregated.
- Offer a visual and detailed summary of opinions.

The steps involved in aspect-based opinion mining are: A) collect, crawl the opinions from the internet, B) Extract the entities that were evaluated, C) extract the aspects of the entity that were commented and how they were described, D) determine whether the opinions about extracted aspects are positive, negative or neutral, E) summarize the results for each aspect. The core tasks are aspect extraction (C) and aspect sentiment classification (D).

There are two types of aspects: explicit and implicit. **Explicit aspects** are aspects that were explicitly mentioned, for example in the sentence *„the battery life is good"* battery life represents an aspect. **Implicit aspects** are those that were not mentioned in text but they are implied from context. In the sentence *„the phone is expensive"*, the aspect that it refers to is *„price"* but is not explicitly mention and thus represents an implicit aspect.

2. Related work

Aspect extraction is performed based on a vocabulary. There are two approaches that were used in creation of the vocabulary: manual or automatic.

Some of related works extract a simple list of features; others organize them into a hierarchy like taxonomies or ontologies. Aspect extraction methods should take in account the fact that people use different word for referring to the same aspect, thus a form of aspect structuring is preferable. Work related to aspect extraction mainly concerned explicit aspects, not much work has been done for implicit aspects.

In [1] the first step of Association rule mining was used to create a list of frequent itemsets with a minimum support of 1%. Two pruning methods were used in to eliminate redundant or unuseful aspects from the candidate frequent aspects. Next opinion words are extracted by extracting the adjacent adjective that modifies the noun/noun phrase that is a frequent aspect. For infrequent aspect extraction for each sentence, if it contained one or more opinion words but no frequent aspect, then the nearest noun/noun phrase of the opinion word was considered an infrequent aspect. The tasks of sentiment classification and summarization were described in a subsequent paper [2]. For sentiment orientation they calculated the sentiment of the sentence containing the aspect [2], thus they did not extract exactly how the aspects were described and then determine the sentiment orientation of the extracted description. The summary format proposed in [2] shows for each aspect, the sentences containing that aspect, grouped by sentiment orientation (positive and negative) , thus is not visual enough, the user still has to read the sentences that mention the aspect. Another disadvantage of this method is that a large number of reviews are needed to construct the aspect list. Also, the similar extracted aspects are not grouped together which causes multiple entries in the summary for same aspect expressed with different aspect expressions.

In [3], [4] a new method called *double propagation* was proposed for extracting both opinion words and aspects simultaneously by exploiting certain syntactic relations of opinion words and aspects due to the fact that opinion words are used to modify aspects. The relations were described using dependency grammar. The method needs only an initial set of opinion word seeds to extract new opinion words and aspect, and the newly extracted terms are used to extract more aspects and opinion words. For example if a known opinion word is in a direct dependency with a noun, then that noun is extracted as an aspect. Polarities are assigned to newly extracted aspects and opinion words, based on some rules. Also three aspect pruning strategies were used to eliminate non aspect terms. In these papers [3], [4] the authors did not focus on summarizing opinions. Although double propagation represents a natural method that gives good results, the disadvantage is that the extracted aspect expressions that refer to the same aspect are not grouped together and thus it would be difficult to create a summary for each aspect.

In [5] an ontology for movie reviews was manually created which was then iteratively expanded with new concepts from sentences that contained a conjunction word and at least one concept seed. An opinion lexicon was created with the help of SentiWordNet. They used the hierarchy structure to calculate the positive, negative and neutral opinion orientation of high-level concepts through the child nodes.

An ontology-based approach for French restaurant reviews was also proposed in [6]. An existing ontology was used which was then translated to French and then adapted. First opinion expressions were extracted based on some rules. Next explicit and implicit aspects are extracted and then Associating opinions expressions to extracted aspects based on some rules. Ontology properties are used in order to retrieve the associated concept in the ontology. Also an automatic expansion of opinion lexicon and a manual expansion of the ontology is performed.

In [5] [6] the summarization task is missing. The advantage of an ontology-based approach is that the aspects are already structured and implicit aspects could be extracted as well as in [6].

For practical applications we think that aspect structuring and summarization are a must. Also the correct approach for sentiment classification will be to first extract how the aspects were described and

then perform sentiment analysis on extracted description, not on the sentence that contains that aspect.

Based on studying related works and their experimental results, we think that the best approach will be to manually create a structured knowledge representation of aspects, like a taxonomy or ontology and use it for extracting aspects that were commented and also for extracting how the aspects were described, based on syntactic relations.

3. Proposed architecture

Customers express their opinions using different words. Using a simple list of aspects for the summarization task will result in duplicate aspects in the summary, which is not desired. We manually created our vocabulary that consists of a list of aspects, each with its own synonyms.

Table 1. Example vocabulary items

Aspect	Synonyms
battery	battery life;talk time;
size	screen size;inch;width;height;diagonal;
memory	RAM;gb;mb;storage;hard disk;
display	screen;graphics;color;
resolution	pixel;camera resolution;
camera	picture;photo;
weight	
sound	audio;speaker;speakerphone;
music	mp3;mp3 player;
video	movie;film;
earphone	headphone;
keyboard	querty;button;keypad;typing;

First we retrieved the products from BestBuy that belong to "*Mobile Phones*" or "*Tablets and iPad*" category, then reviews are retrieved from BestBuy and filtered, we were interested only reviews that belong to products stored in our database.

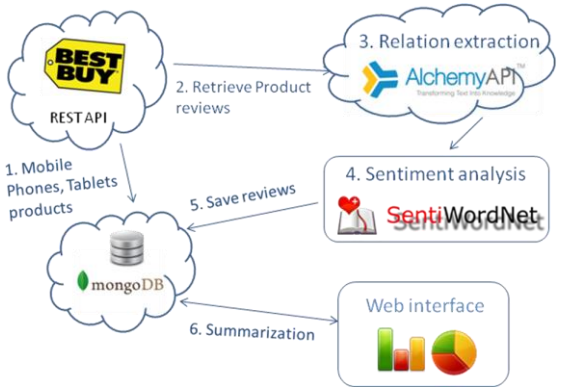

Fig. 1 - System overview

From each review title and content *Subject-Action-Object* relations were extracted using Alchemy API Relation Extraction Service [7]. For example, for the sentence „*The battery life is great*." Alchemy API will extract the following relation: Subject: *The battery life* - Action*: is* - Object: *great.* Some of the extracted relations are not referring to an aspect and since we are only interested in relations about a product aspect, a filtering is needed. The filtering of relations is based on the vocabulary and we considered the Subject of the relation an aspect and for description of the aspect the Action and Object of the relation were used.

Next for each filtered relation sentiment analysis was performed on the aspect description (action + object of the relation), not on the sentence that contained the aspect. The algorithm used was based on SentiWordNet [8].

For data storage we used MongoDB, a NoSQL document oriented database.

Last step is summarization of opinions. For a product name entered by user in our web interface, the system retrieves the extracted relations and their sentiment and creates a summary per extracted aspect as follows:

- A stacked column chart in which each column corresponds to an extracted aspect and the color of the stacks represent the number of positive descriptions in green, neutral descriptions in yellow and
- A pie chart for each aspect, with three slices: green for the number of positive descriptions, yellow for neutral and red for negative descriptions.

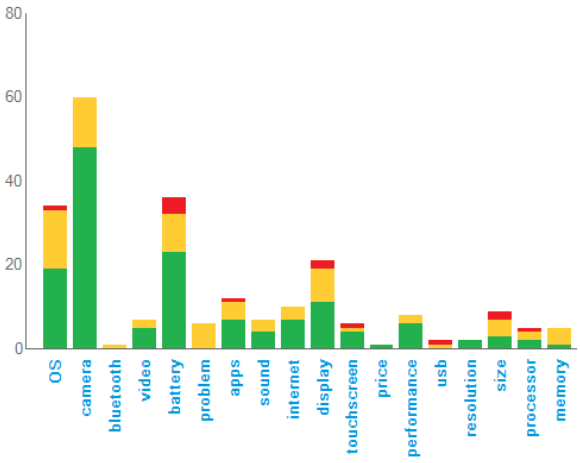

Fig. 2 - Summary for "iphone 4s"

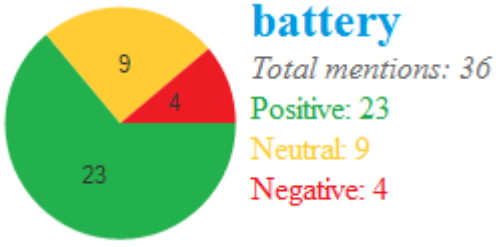

Fig. 3 - Summary of "battery" aspect for "iphone 4s" product

When a user clicks on an aspect name, a popup containing the aspect descriptions is shown, each

descriptions colored according to its sentiment orientation as shown in Fig. 4.

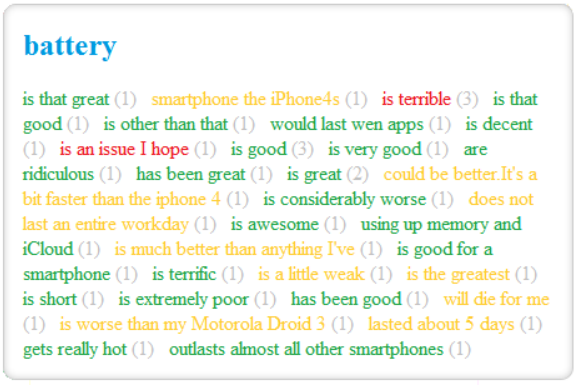

Fig. 4 - Extracted battery descriptions for "iphone 4s" product

Also we provide a product opinion comparison. For product names separated by ";" the system retrieves the extracted relations for each product and a score is given for each product aspect. An example of comparison is shown in Fig. 5.

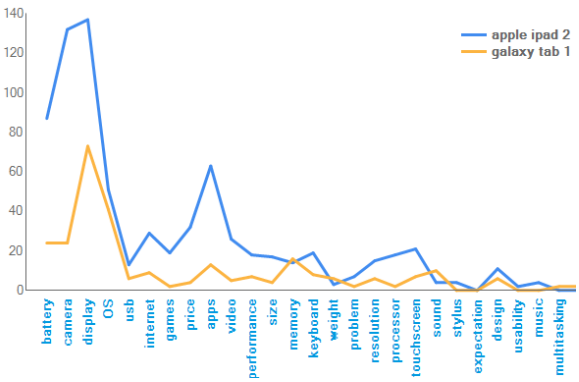

Fig. 5 - Comparison of "apple ipad 2" and "galaxy tab 1"

4. Discussion

A weakness of our approach is that *Subject-Action-Object* relations extracted by Alchemy API Relation Extraction [7] are more suited for facts instead of opinions and direct dependencies are not extracted. For example in case of „ *It has great battery life*", *"great battery life"* will be extracted as the *Object* of the relation and it will not be taken in account since we are only looking for aspects in the *Subject* of the relation. In this case a subsequent text analysis is needed to extract the aspect and the description from the *Object* of the relation.

Another weakness is that for product comparison we used a simple formula to calculate a score for each aspect based on the number of mentions and we are still looking for a more suitable one.

5. Conclusions

Our approach shows that is possible to create a system for aspect-based opinion mining with real world application.

Further work should be to implement our own aspect extraction algorithm using the current vocabulary or develop a domain ontology and exploit syntactic relations between aspects and opinion words.

Also a future work would be adding the possibility that a user can specify the aspects he/she is interested, and only show summary for that aspects, if extracted.

Since we currently only analyse reviews for mobile phones and tablets from BestBuy, another further step could be to retrieve reviews from other Internet sources like Twitter, eBay, Amazon or blogs and expand our system for other product categories as well.

References

[1] M. Hu and B. Liu, "Mining opinion features in customer reviews," in *Proceedings of the 19th National Conference on Artificial Intelligence (AAAI'04)*, 2004, pp. 755–760.

[2] M. Hu and B. Liu, "Mining and Summarizing Customer Reviews," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '04)*, 2004.

[3] G. Qiu, B. Liu, J. Bu, and C. Chen, "Expanding Domain Sentiment Lexicon through Double Propagation," in *Proceedings of the 21st international jont conference on Artifical intelligence (IJCAI-09)*, 2009, pp. 1199-1204.

[4] G. Qiu, B. Liu, J. Bu, and C. Chen, "Opinion word expansion and target extraction through double propagation," *Computational Linguistics*, vol. 37, no. 1, 2011.

[5] L. Zhao and C. Li, "Ontology Based Opinion Mining for Movie Reviews," in *Proceedings of the 3rd International Conference on Knowledge Science, Engineering and Management*, 2009, pp. 204-214.

[6] A. Cadilhac, F. Benamara, and N. Aussenac-gilles, "Ontolexical resources for feature based opinion mining□: a case-study," in *Proceedings of the 6thWorkshop on Ontologies and Lexical Resources (Ontolex2010)*, 2010, no. August, pp. 77-86.

[7] "Alchemy API Relation Extraction Service." [Online]. Available: http://www.alchemyapi.com/api/relation/.

[8] A. Esuli and F. Sebastiani, "SENTIWORDNET□: A Publicly Available Lexical Resource for Opinion Mining," pp. 417-422, 2006.