ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 19, April 2016

MINING ONLINE PRODUCT REVIEWS

[1]ANRUPRIYA V,[2] HARSHINI N,[3]Mr. K.RAJAGANAPTHY

[1,2] UG Scholars,[3] ASSOCIATE PROFESSOR
KCG COLLEGE OF TECHNOLOGY

Abstract – Social media is emerging rapidly on the internet. This media knowledge helps people, company and organizations to analyze information for important decision making. Opinion mining is also called as sentiment analysis which involves in building a system to gather and examine opinions about the product made in reviews or tweets, comments, blog posts on the web. Sentiment is classified automatically for important applications such as opinion mining and summarization. To make valuable decisions in marketing analysis where implement sentiment classification efficiently. Reviews contain sentiment which is expressed in a different way in different domains and it is costly to annotate data for each new domain. The analysis of online customer reviews in which firms cannot discover what exactly people liked and did not like in document-level and sentence-level opinion mining. The proposed system is based on phrase-level to examine customer reviews. Phrase-level opinion mining is also well-known as aspect based opinion mining. It is used to extract most important aspects of an item and to predict the orientation of each aspect from the item reviews. The projected system implements aspect extraction using frequent itemset mining in customer product reviews and mining opinions whether it is positive or negative opinion. It identifies sentiment orientation of each aspect by supervised learning algorithms in customer reviews. By using the information obtained from customer reviews, model the relationships among products by constructing a Decision tree. Then mine this tree to determine the relative quality of products. This is implemented by comparing the products in multiple e-commerce websites and generates a comparative result.

1. INTRODUCTION

With the rapid expansion of e-commerce, more and more products are sold on the Web, and more and more people are also buying products online. In order to enhance customer satisfaction and shopping experience, it has become a common practice for online merchants to enable their customers to review or to express opinions on the products that they have purchased. With more and more common users becoming comfortable with the Web, an increasing number of people are writing reviews. As a result, the number of reviews that a product receives grows rapidly. Some popular products can get hundreds of reviews at some large merchant sites. Furthermore, many reviews are long and have only a few sentences containing opinions on the product. This makes it hard for a potential customer to read them to make an informed decision on whether to purchase the product. If he/she only reads a few reviews, he/she may get a biased view. The large number of reviews also makes it hard for product manufacturers to keep track of customer opinions of their products. For a product manufacturer, there are additional difficulties because many merchant sites may sell its products, and the manufacturer may (almost always) produce many kinds of products. In this research, we study the problem of generating feature-based summaries of customer reviews of products sold online. Here, features broadly mean product features (or attributes) and functions. Given a set of customer reviews of a particular product, the task involves three subtasks: (1) identifying features of the product that customers have expressed their opinions on (called product features); (2) for each feature, identifying review sentences that give positive or negative opinions; and (3) producing a summary using the discovered information. Let us use an example to illustrate a feature-based summary. Assume that we summarize the reviews of a particular digital camera, digital_camera_1. The summary looks like the following:

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

As indicated above, our task is performed in three main steps:

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 19, April 2016

(1) Mining product features that have been commented on by customers. We make use of both data mining and natural language processing techniques to perform this task. However, for completeness, we will summarize its techniques in this paper and also present a comparative evaluation.

(2) Identifying opinion sentences in each review and deciding 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 (whether the opinion expressed in the sentence is positive or negative), we perform three subtasks. First, a set of adjective words (which are normally used to express opinions) is identified using a natural language processing method. These words are also called opinion words in this paper. Second, for each opinion word, we determine its semantic orientation, e.g., positive or negative. A bootstrapping technique is proposed to perform this task using WordNet. Finally, we decide the opinion orientation of each sentence. An effective algorithm is also given for this purpose.

(3) Summarizing the results. This step aggregates the results of previous steps and presents them in the format of Figure 1. Section 3 presents the detailed techniques for performing these tasks. A system, called FBS (Feature-Based Summarization), has also been implemented.

2. RELATED WORK

The authors of [3] describes summarize the work of sentiment analysis and polarity shift, and then review the technique of data expansion.

•Sentiment Analysis and Polarity Shift

According to the levels of granularity, tasks in sentiment analysis can be divided into four categorizations: document level, sentence-level, phrase-level, and aspect-level sentiment analysis. Focusing on the phrase/sub sentence and aspect-level sentiment analysis, with a lexicon of words with established prior polarities, and identify the "contextual polarity" of phrases, based on some refined annotations. Further the authors of [2][1] combined different kinds of negators with lexical polarity items though various compositional semantic models, both heuristic and machine learned, to improved sub sentential sentiment analysis. The authors of [3] developed a semi-supervised model for sub sentential sentiment analysis that predicts polarity based on the interactions between nodes independency graphs, which potentially can induce the scope of negation. In aspect-level sentiment analysis, the polarity shift problem was considered in both corpus- and lexicon based methods. For document- and sentence-level sentiment classification, there are two main types of methods in the literature: Term-counting and machine learning methods. In Term-counting methods, the overall orientation of a text is obtained by summing up the orientation scores of content words in the text, based on manually-collected or external lexical resources. In machine learning methods, sentiment classification is regarded as a statistical classification problem, where a text is represented by bag-of words; then, the supervised machine learning algorithms are applied as classifier [4] Accordingly, the way to handle polarity shift also differs in the two types of methods. The term-counting methods can be easily modified to include polarity shift. One common way is to directly reverse the sentiment of polarity-shifted words, and then sum up the sentiment score word by word. Compared with term counting methods, the machine learning methods are more widely discussed in the sentiment classification literatures. However, it is relatively hard to integrate the polarity shift information into the BOW model in such methods. For example, in paper [1] proposed a method by simply attaching "NOT" to words in the scope of negation, so that in the text "I don't like book", the word "like" becomes a new word "like NOT". This was reported that this method only has slightly negligible effects on improving the sentiment classification accuracy. There were also some attempts to model polarity shift by using more linguistic features or lexical resources. For example, proposed to model negation by looking for specific part-of-speech tag patterns. Proposed to use syntactic parsing to capture three types of valence shifter. Their results showed that handling polarity shift improves the performance of term-counting systems significantly, but the improvements upon the baselines of machine learning systems are very slight (less than 1 percent). Proposed a machine learning method based on a lexical dictionary extracted from General Inquirer1 to model polarity-shifters for both word-wise and sentence-wise sentiment classification. There were still some approaches that addressed polarity shift without complex linguistic analysis and extra annotations. Classification models are then trained based on each of the two parts. An ensemble of two component classifiers is used to provide the final polarity of the whole text.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 19, April 2016

Table 2.1 An example of creating Reversed training reviews

	Review Text	Class
Original review	I don't like this book. It is boring.	Negative
Reversed review	I like this book. It is interesting.	Positive

- Data expansion by creating reversed reviews Based on an antonym dictionary, for each original review, the reversed review is created according to the following rules:

- Text reversion. If there is a negation, the scope of negation is detected. All sentiment words out of the scope of negation are reversed to their antonyms. In the scope of negation, negation words (e.g., "no", "not", "don't", etc.) are removed, but the sentiment words are not reversed;

- Label reversion. For each of the training review, the class label is also reversed to its opposite (i.e., positive to negative, or vice versa), as the class label of the reversed review.

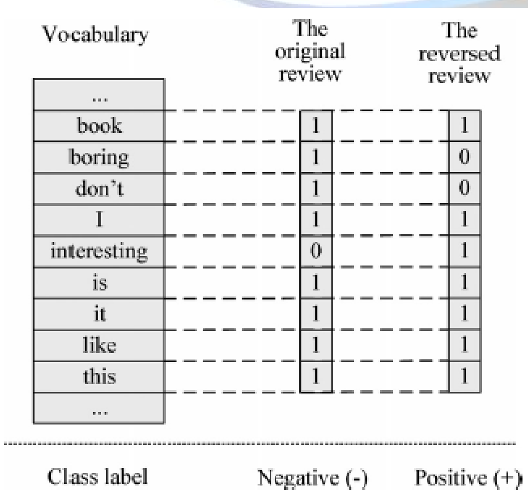


Figure 2.1 The BOW representations of the original and the reversed reviews in Table 2.1

The authors of paper [2] give a detail description on how to rank the products based on user reviews. The authors proposed a model consisting of three stages to enhance the review reliability to the product evaluation. The first stage filters out the sentences that contain comments which are unrelated to the product quality. The second stage derives weights for a review based on its helpfulness votes and age, i.e. since the posting date. The third stage calculates the product's overall ranking score. In their ranking system, the ranking score is determined by the review contents, relevance of a review to the product quality, helpful votes and total votes from posterior customers, posting date and durability of reviews.

Three types of features are used to the classification task:

- *Brand-level (PL) feature* includes the product brand names, e.g. Nikon or Canon, and the model names (e.g. 550d or D90). This feature counts the number of lexical matches in each sentence.

- *Semantic-level (SL) feature* are the subjective and objective words (positive or negative) describing products.

- *Product-level (FL) feature* is the number of product specification attributes, such as camera pixels and lens, mentioned in the sentences and the number of words related to the customer services, such as shipping and customer support

Filtering Mechanism

A relevant sentence is either an overall or feature-based comment on a product. It evaluates at least one aspect of a product and provides convincing opinions. The most commonly seen irrelevant review sentence is about the customer service of the seller. In this study, consider the task of differentiating the irrelevant sentences a binary classification problem. The authors use a Supporting Vector Machine (SVM), a well-known supervised machine learning algorithm, to train a hypothesis function, *h*. Each review sentence becomes trained data in the form of *f sentence; h*(*sentence*)*g* pair. Given a review sentence, first construct its feature vector **X**, which is fed to the

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 19, April 2016

SVM to generate a relevance score based on the linear regression model:

$h(\mathbf{X}) = B^T \mathbf{X} + b$ (2.1)

Where B is the coefficient vector of weights and b is the intercept. The training set contains 1000 sentences collected manually. The value of $h(\mathbf{X})$ indicates the probability that the sentence is relevant or not. The details of extracting feature vectors are given in the next section.

Three types of features are used to the classification task:

- *Brand-level (PL) feature* includes the product brand names, e.g. Nikon or Canon, and the model names (e.g. 550d or D90). This feature counts the number of lexical matches in each sentence.
- *Semantic-level (SL) feature* are the subjective and objective words (positive or negative) describing products.
- *Product-level (FL) feature* is the number of product specification attributes, such as camera pixels and lens, mentioned in the sentences and the number of words related to the customer services, such as shipping.

Helpfulness Vote

The perceived importance of a review depends on not only its quality but also the helpfulness votes cast by the posterior customers. The helpfulness of a review is determined by the total number of helpful votes and the number of total votes.

Age of Review and Durability

In this model, reviews posted more recently receive higher weights in assessing their importance. Assume newer reviews are likely written by customers who read some of the reviews posted earlier. These reviews shall receive higher credits because they are more influential to potential customers. Without adding weights to the newer reviews, they would contribute less to the ranking score, as they are "young" and likely receive fewer votes. Another important consideration is newer releases/versions of the same products. Correspondingly, the number of reviews for a product version released earlier is likely higher than the product version released recently. In order to balance the contributions to the ranking scores among the similar products and minimize the effects from large volumes gaps, reduce the importance of older reviews and increase the weight for newer reviews. Figure 4 shows the growing trends of the review numbers during a 13-week period since the product release. The following exponential equation 3 is used to model the age importance of a review.

$T(r, p) = e^{B (tr-t0)} + d$ (2.2)

Where $T(r,p)$ is the estimated weight, $t0$ is the product p release date, tr is the published date of review r, B controls the decay rate of $T(r; p)$, and d is an initializing factor. Note that B and d have different values when calculating the age weights for products from different categories.

Sentence Splitter and Part-Of-Speech Tagging

A customer review typically consists of several sentences. It is not uncommon to see multiple positive and negative opinions of a product in a single review. For example, a review of a digital camera may use a few sentences to praise the picture quality and others to criticize the weight and colour of the camera. It is not easy to determine the sentiment orientation of such a review as a whole. To simplify this problem, in this paper they have split reviews into sentences. The sentences are then assigned positive or negative sentiments. In this study, do not consider sentences expressing both positive and negative sentiments. MXTERMINATOR is used to split reviews into sentences. Most sentiment bearing words are adjectives. These adjectives determine if a sentence is subjective, and whether it expresses positive or negative sentiment or both. In order to help to identify the sentiment orientation of a sentence, the part-of-speech information is used.

Score Function

A product ranking score is calculated based on the review contents by incorporating two factors: helpfulness vote and review age. For reviews containing more than one relevant sentence label each with a positive, negative, or neutral tag. The difference between the number of positive and negative sentences determines the polarity of a review. The weights derived from the helpfulness vote and review age are then applied to the ranking score calculation. The ranking score of a product is the sum of all weighted scores of individual reviews. The following equation computes the ranking score S of product p.

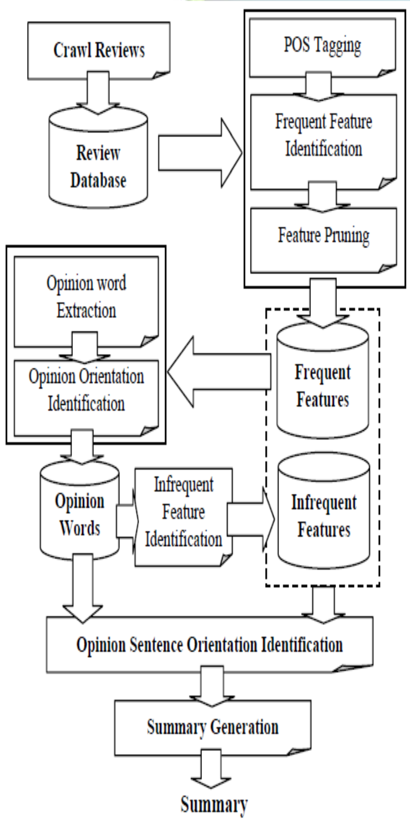
ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 19, April 2016

$$S(p) = \frac{\sum_{all\ r} Polarity(r,p) \cdot T(r,p) \cdot H(r,p)}{\sum_{all\ r} H(r,p) \cdot \sum_{all\ r} T(r,p)} \quad (2.3)$$

Where *Polarity*(*r; p*) = *Pos*(*r, p*) - *Neg*(*r; p*). *Pos*(*r; p*) and *Neg*(*r; p*) are the numbers of positive and negative sentences in review *r* of product *p*, respectively.

3. THE PROPOSED TECHNIQUES

Figure 1 gives the architectural overview of our opinion summarization system.



The inputs to the system are a product name and an entry Web page for all the reviews of the product. The output is the summary of the reviews as the one shown in the introduction section. The system performs the summarization in three main steps (as discussed before): (1) mining product features that have been commented on by customers; (2) identifying opinion sentences in each review and deciding whether each opinion sentence is positive or negative; (3) summarizing the results. These steps are performed in multiple sub-steps.

Given the inputs, the system first downloads (or crawls) all the reviews, and put them in the review database. It then finds those “hot” (or frequent) features that many people have expressed their opinions on. After that, the opinion words are extracted using the resulting frequent features, and semantic orientations of the opinion words are identified with the help of WordNet. Using the extracted opinion words, the system then finds those infrequent features. In the last two steps, the orientation of each opinion sentence is identified and a final summary is produced. Note that POS tagging is the part-of-speech tagging from natural language processing, which helps us to find opinion features. Below, we discuss each of the sub-steps in turn

3.1 Part-of-Speech Tagging (POS)

Product features are usually nouns or noun phrases in review sentences. Thus the part-of-speech tagging is crucial. We used the NLProcessor linguistic parser [31] to parse each review to split text into sentences and to produce the part-of-speech tag for each word (whether the word is a noun, verb, adjective, etc.). The process also identifies simple noun and verb groups (syntactic chunking). The following shows a sentence with POS tags.

<S><NG><W C='PRP' L='SS' T='w' S='Y'> I </W></NG>
<VG><W C='VBP'> am </W><W C='RB'> absolutely </W></VG>
<W C='IN'> in </W><NG><W C='NN'> awe </W></NG><W C='IN'> of </W>
<NG><W C='DT'> this </W><W C='NN'> camera </W></NG><W C='.'> . </W></S>

NLProcessor generates XML output. For instance, <W C=‘NN’> indicates a noun and <NG> indicates a noun group/noun phrase. Each sentence is saved in the review database along with the POS tag information of each word in the sentence. A transaction file is then created for the generation of frequent features.

3.2 Frequent Features Identification

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 19, April 2016

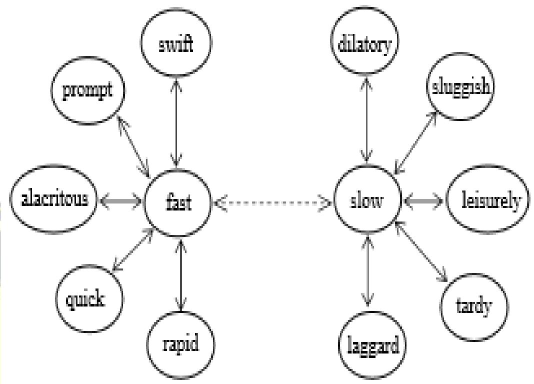
This sub-step identifies product features on which many people have expressed their opinions. Before discussing frequent feature identification, we first give some example sentences from some reviews to describe what kinds of opinions that we will be handling. Since our system aims to find what people like and dislike about a given product, how to find the product features that people talk about is the crucial step. However, due to the difficulty of natural language understanding, some types of sentences are hard to deal with. Let us see an easy and a hard sentence from the reviews of a digital camera: "The pictures are very clear." In this sentence, the user is satisfied with the picture quality of the camera, picture is the feature that the user talks about. While the feature of this sentence is explicitly mentioned in the sentence, some features are implicit and hard to find. For example, "While light, it will not easily fit in pockets." This customer is talking about the size of the camera, but the word size does not appear in the sentence. In this work, we focus on finding features that appear explicitly as nouns or noun phrases inthe reviews. We leave finding implicit features to our future work. Here, we focus on finding frequent features, i.e., those features that are talked about by many customers.

3.3 Opinion Words Extraction

We now identify opinion words. These are words that are primarily used to express subjective opinions. If a sentence contains one or more product features and one or more opinion words, then the sentence is called an opinion sentence. We extract opinion words in the following manner (Figure 3): for each sentence in the review database if (it contains a frequent feature, extract all the adjective words as opinion words
) for each feature in the sentence the nearby adjective is recorded as its effective opinion. /* A nearby adjective refers to the adjacent adjective that modifies the noun/noun phrase that is a frequent feature. */ Figure 3: Opinion word extraction.

3.4 Orientation Identification for Opinion Words

For each opinion word, we need to identify its semantic orientation, which will be used to predict the semantic orientation of each opinion sentence. The semantic orientation of a word indicates the direction that the word deviates from the norm for its semantic group. Words that encode a desirable state (e.g., beautiful, awesome) have a positive orientation, while words that represent undesirable states have a negative orientation (e.g., disappointing).



3.5 Summary Generation

After all the previous steps, we are ready to generate the final feature-based review summary, which is straightforward and consists of the following steps:

- For each discovered feature, related opinion sentences are put into positive and negative categories according to the opinion sentences' orientations. A count is computed to show how many reviews give positive/negative opinions to the feature.
- All features are ranked according to the frequency of their appearances in the reviews. Feature phrases appear before single word features as phrases normally are more interesting to users. Other types of rankings are also possible. For example, we can also rank features according the number of reviews that express positive or negative opinions. The following shows an example summary for the feature "picture" of a digital camera. Note that the individual opinion sentences (and their corresponding reviews, which are not shown here) can be hidden using a hyperlink in order to enable the user to see a global view of the summary easily.

Feature: picture Positive: 12 • Overall this is a good camera with a really good picture clarity. • The pictures are absolutely amazing - the camera captures the minutest of details. • After nearly 800 pictures I have found that this camera takes incredible pictures. … Negative: 2 • The pictures come out hazy if your hands shake even for a moment during the entire process of taking a picture. • Focusing on a display rack about 20 feet away in a brightly lit room during day time,

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 19, April 2016

pictures produced by this camera were blurry and in a shade of orange.

3.5 Infrequent Feature Identification

Frequent features are the "hot" features that people comment most about the given product. However, there are some features that only a small number of people talked about. These features can also be interesting to some potential customers and the manufacturer of the product. The question is how to extract these infrequent features? Considering the following sentences:

"The pictures are absolutely amazing."
"The software that comes with it is amazing."

Sentences 1 and 2 share the same opinion word amazing yet describing different features: sentence 1 is about the pictures, and sentence 2 is about the *software*. Since one adjective word can be used to describe different objects, the opinion words are used to look for features that cannot be found in the frequent feature generation step using association mining.

The nearest noun/noun phrase as the noun/noun phrase that the opinion word modifies because that is what happens most of the time. This simple heuristic seems to work well in practice.

A problem with the infrequent feature identification using opinion words is that it could also find nouns/noun phrases that are irrelevant to the given product. The reason for this is that people can use common adjectives to describe a lot of objects, including both interesting features that is wanted and irrelevant ones. This, however, is not a serious problem because the number of infrequent features, compared with the number of frequent features, is small. They account for around 15-20% of the total number of features as obtained in our experimental results.

Infrequent features are generated for completeness. Moreover, frequent features are more important than infrequent ones. Since we rank features according to their p-supports, those wrong infrequent features will be ranked very low and thus will not affect most of the users.

3.6 Predicting The Orientations Of Opinion Sentences

This is the step of predicting the orientation of an opinion sentence, i.e., positive or negative. In general, the dominant orientation is used for the opinion words in the sentence to determine the orientation of the sentence. That is, if positive/negative opinion prevails, the opinion sentence is regarded as a positive/negative one. In the case where there is the same number of positive and negative opinion words in the sentence, predicts the orientation using the average orientation of effective opinions or the orientation of the previous opinion sentence.

Predicting the semantic orientation of an opinion sentence:

1. The user likes or dislikes most or all the features in one sentence. The opinion words are mostly either positive or negative, e.g., there are two positive opinion words, good and exceptional in "overall this is a good camera with a really good picture clarity & an exceptional close-up shooting capability."

2. The user likes or dislikes most of the features in one sentence, but there is an equal number of positive and negative opinion words, e.g., "the auto and manual along with movie modes are very easy to use, but the software is not intuitive." There is one positive opinion *easy* and one negative opinion *not intuitive*, although the user likes two features and dislikes one.

3. All the other cases.

For case 1, the dominant orientation can be easily identified. This is the most common case when people express their opinions. For case 2, use the average orientation of effective opinions of features. Effective opinion is assumed to be the most related opinion for a feature. For case 3, set the orientation of the opinion sentence to be the same as the orientation of previous opinion sentence. Use the context information to predict the sentence orientation because in most cases, people express their positive/negative opinions together in one text segment, i.e., a few consecutive sentences.

For a sentence that contains a but clause (sub-sentence that starts with but, however, etc.), which indicates sentimental change for the features in the clause, first use the effective opinion in the clause to decide the orientation of the features. If no opinion appears in the clause, the opposite orientation of the sentence will be used.

3.7 Summary Generation

After all the previous steps, this is ready to generate the final feature-based review summary, which is straightforward and consists of the following steps:

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 19, April 2016

- For each discovered feature, related opinion sentences are put into positive and negative categories according to the opinion sentences' orientations. A count is computed to show how many reviews give positive/negative opinions to the feature.
- All features are ranked according to the frequency of their appearances in the reviews. Feature phrases appear before single word features as phrases normally are more interesting to users. Other types of rankings are also possible. For example, this can also rank features according the number of reviews that express positive or negative opinions.

The following shows an example summary for the feature "*picture*" of a digital camera. Note that the individual opinion sentences (and their corresponding reviews, which are not shown here) can be hidden using a hyperlink in order to enable the user to see a global view of the summary easily.

Feature: **picture**
Positive: 12

- Overall this is a good camera with a really good picture clarity.
- The pictures are absolutely amazing - the camera captures the minutes of details.
- After nearly 800 pictures I have found that this camera takes incredible pictures.

…
Negative: 2

- The pictures come out hazy if your hands shake even for a moment during the entire process of taking a picture.
- Focusing on a display rack about 20 feet away in a brightly lit room during day time, pictures produced by this camera were blurry and in a shade of orange.

4. CONCLUSION

In this paper, we proposed a set of techniques for mining and summarizing product reviews based on data mining and natural language processing methods. The objective is to provide a feature-based summary of a large number of customer reviews of a product sold online. Our experimental results indicate that the proposed techniques are very promising in performing their tasks. We believe that this problem will become increasingly important as more people are buying and expressing their opinions on the Web. Summarizing the reviews is not only useful to common shoppers, but also crucial to product manufacturers.

5. REFERENCES

[1] A.Jeyapriya, C.S.Kanimozhi Selvi (2015) 'Extracting Aspects and Mining Opinions in Product Reviews using Supervised Learning Algorithm' IEEE SPONSORED 2ND INTERNATIONAL CONFERENCE ON ELECTRONICS AND COMMUNICATION SYSTEMS(ICECS '2015)

[2] Kunpeng Zhang, Yu Cheng, Wei-keng Liao, Alok Choudhary (2015) 'Mining Millions of Reviews: A Technique to Rank Products Based on Importance of Reviews' - EECS department, Northwestern University

[3] Rui Xia, Feng Xu, Chengqing Zong, Qianmu Li, Yong Qi, and Tao Li (2015) 'Dual Sentiment Analysis: Considering Two Sides of One Review' - IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 27, NO. 8, AUGUST 2015

[4] X. Yu, Y. Liu, X. Huang, and A. An (2012) 'Mining online reviews for predicting sales performance: A case study in the movie domain' - IEEE Trans. Knowl. Data Eng., vol. 24, no. 4, pp. 720–734, Apr. 2012.