

Statistical Features Identification for Sentiment Analysis using Machine Learning Techniques

Ahmad Kamal

Department of Mathematics
Jamia Millia Islamia (A Central University)
New Delhi - 110025, India
E-mail: ah.kamal786@gmail.com

Muhammad Abulaish[#], SMIEEE

Department of Computer Science
Jamia Millia Islamia (A Central University)
New Delhi - 110025, India
E-mail: mAbulaish@jmi.ac.in

Abstract – Due to increasing fascinating trend of using internet and online social media, user-generated contents are growing exponentially on the Web, containing users' opinion on various products. In this paper, we have proposed a sentiment analysis system which combines rule-based and machine learning approaches to identify feature-opinion pairs and their polarity. The efficiency of the proposed system is established through experimentation over customer reviews on different electronic products.

Keywords – Opinion mining, Sentiment analysis, Text Mining, Feature identification, Machine learning.

I. INTRODUCTION

Web opinion sources such as merchant sites, Web forums, discussion groups, and blogs are rapidly emerging containing precious information useful for both customers and manufacturers. Due to easy accessibility of the Web, opinion sources are used as a platform by individual users to share their experiences or opinions. Users' opinions play an important role both for customers as well as for manufacturers to take proper decision. The feedbacks of existing users are useful for new customers in choosing a right product, whereas on the other hand, it helps product manufacturers to know the strength and weaknesses of their products from the perspective of end-users. Such opinion is very informative in developing marketing and product development plans.

Recently, feature-based opinion mining techniques are gaining momentum in which every granule of customer reviews is processed to identify product features and users' opinions expressed over them. A study in [1] reveals that a complete opinion along with its relevant feature is always expressed in one sentence. If a sentence contains product feature, its nearest opinion word can be associated using Parts-Of-Speech (POS) information and dependency relationship. But, a large number of product features present as noun phrases in review sentences are referenced by anaphoric pronouns present in succeeding sentences of a review document. Anaphora resolution is one of the classical computational linguistics problems, and its resolution is important for feature-based opinion mining

research, otherwise many opinion-related information will be left unnoticed.

In addition to the identification of features and opinions, another challenge related to the development of an effective sentiment analysis system is to classify sentiment or polarity of opinion bearing words. Sentiment analysis task is context sensitive and domain dependent despite of the fact that general view of positive and negative opinions remains consistent across different domains [2]. Also, presence of negation word such as *not* in a review sentence, makes the task of sentiment analysis difficult, as sentiment of the related opinionated word requires inference in opposite sense for proper polarity determination.

Our aim in this work is to propose the design of a feature-based sentiment analysis system at the intersection of rule-based and machine learning approaches. The rule-based approach is exploited to identify feasible feature-opinion pairs from opinion sources, whereas the machine learning approach is used to determine the polarity of opinion bearing words. The rule-based approach is augmented with anaphora resolution mechanism which uses backtracking to identify features expressed using pronouns. Initially, information components are extracted to fill a template $\langle f, m, o \rangle$, where f represents a product feature, o represents an opinion expressed over f , and m is a modifier used to model the degree of expressiveness of o . Further, word-level sentiment classification method is implemented with the aid of statistical approach and supervised machine learning technique to determine the polarity of opinionated words.

The remaining paper is structured as follows. Section 2 presents a brief review of the existing opinion mining and sentiment analysis techniques. Section 3 presents the functional detail of the proposed system. The experimental setup and results are presented in section 4. Finally, section 5 concludes the paper.

II. RELATED WORK

Research efforts in the field of feature-based opinion mining associate product features and opinion words on the basis of their co-occurrence at sentence level [1,3]. In [3], a semi-supervised technique, termed as double propagation, is proposed to extract features and opinion words using seed opinion lexicon. Further, extracted features and opinions are exploited for identifying new features and opinions. Since a

[#] To whom correspondence should be addressed

complete opinion along with its relevant feature is always expressed in one sentence [1], feature and opinion pair extraction can be performed at sentence-level to avoid their false associations. Though research efforts have not been applied on a large-scale for binding product features expressed through anaphoric pronouns, the authors in [4] proposed a supervised learning method for anaphora resolution through decomposing the problem into two sub-tasks - anaphoric determination and antecedent identification. In [5], Jakob and Gurevych presented an unsupervised anaphora resolution approach for feature identification in movie reviews using a rule-based pronominal resolution algorithm, CogNIAC, which resolves only those anaphora satisfying very high confidence-level rules and ambiguous antecedents are left unresolved to achieve high precision. Apart from product feature identification, opinion mining research requires sentiment classification of opinion bearing words. Pang et al. [6] proposed a supervised learning technique for sentiment classification, which is based on unigram model. In addition, some researchers have proposed corpus- and dictionary-based approaches for word-level sentiment classification [7]. Though machine-learning approaches have been proved very effective for classification tasks, they need to be explored on review documents for sentiment classification problem. This paper presents a good mix of rule-based and machine learning approach to identify features, opinions and their polarities.

III. PROPOSED FEATURE-BASED SENTIMENT ANALYSIS SYSTEM

In this section, we present the functional detail of the proposed feature-based opinion mining and sentiment analysis system to identify *feature-opinion* pairs and to determine the sentiment of the opinion bearing words. The proposed system consists of five different functional components - *feature and opinion extraction, feature-opinion binding through anaphora resolution, feasibility analysis, statistical features identification, and sentiment determination using machine learning techniques*. Further details about these components are presented in the following sub-sections.

A. Feature and Opinion Extraction

The input to this component is POS-tagged review documents along with dependency relationships information between the words. We have defined a set of rules to access different types of sentence structures for identification of information components embedded within them. As an enhancement to our rule-set reported in [8], some new rules are defined in this work to tackle the peculiarity and complexity of review documents. Consider the following review sentences:

- The *picture quality* is really *nice, amazing* and *awesome*.
- *Samsung S5830* has a *powerful battery*.

In first sentence, the bigram *picture quality* is a product feature and can be identified using "NN" tag, whereas the word *quality* is related to an adjective *nice* with "NSUBJ" relation. Thus, *nice* can be identified as an opinion. Here, "NN" is a noun compound modifier and "NSUBJ" is a dependency relation used in the Stanford Parser. Further, multiple opinion words *amazing* and *awesome* present in it can be extracted using one or more occurrence of *and* relationship with the opinion word *nice*. As observed in [3], existing features can also be used to identify new feature words. In the second sentence mentioned above, the word *S5830* of the product *Samsung S5830* is the nominal subject of the verb *has* and the word *battery* is the direct object of it. Thus, *battery* can be identified as a new feature word. Further, "AMOD" relationship can be used to identify *powerful* as an opinion word. Based on these observations, a set of rules is designed to identify *feature-opinion* pairs.

B. Feature-Opinion Binding through Anaphora Resolution using Backtracking

In order to enhance the *recall* value of the proposed method, pronominal coreference is performed to identify product features that are referred through anaphora pronouns. For a review sentence, containing opinionated word along with anaphora pronoun, context of zero or more preceding sentences are analyzed to bind best antecedent (product feature) with an anaphora. We have implemented information component extraction mechanism using backtracking in which review documents are accessed in reverse order, i.e., the last sentence containing opinionated word and the anaphora is selected first for binding anaphora with antecedents. All anaphora pronouns present in a sentence that require mapping are extracted, and a set of anaphora $A = \{a_1, a_2, a_3, \dots, a_n\}$ is compiled for proper context determination. For each anaphora $a_i \in A$, proper context is determined to compile a set $N = \{n_{i1}, n_{i2}, n_{i3}, \dots, n_{ik}\}$ consisting of candidate antecedents. The best antecedent $n_{ij} \in N$ as well as $n_{ij} \in F$ is selected for binding with a_i on the basis of specific evaluation criteria that consist of semantic association and recency factor, i.e., antecedent that appear close to the pronoun to be resolved is selected for binding. The backtracking process terminates when beginning of the review document, i.e., first sentence is reached and accessed.

C. Feasibility Analysis

During the information component extraction phase, numerous nouns, verbs and adjectives are extracted that are not relevant to our task. Sometimes, it is observed that verbs are considered as nouns due to parsing error. Therefore, in line with [8], we have handled noisy extraction by calculating reliability score, r_{ij} , for every candidate *feature-opinion* pair (f_i, o_j) , and normalizing this score using *min-max* normalization to fit into the scale [0, 1] as shown in equation 1, where $HS_{(p_{ij})}^n$ denotes hub score of p_{ij} after n^{th} iteration (after convergence) and *NewMax* and *NewMin* values are set to 1 and 0 respectively. This metric

determines the reliability of an opinion expressed over a product feature. Further details about $HS^n_{(p_{ij})}$ can be found in our previous work [8].

$$r_{ij} = \frac{HS^n_{(p_{ij})} - \min_{xy} \{HS^n_{(p_{xy})}\}}{\max_{xy} \{HS^n_{(p_{xy})}\} - \min_{xy} \{HS^n_{(p_{xy})}\}} \times (NewMax - NewMin) + NewMin \quad (1)$$

D. Statistical Features Identification

A statistical learning approach is proposed, and popular association measuring methods, including Pointwise Mutual Information (PMI), Mutual Information (MI), Chi-square, and Log Likelihood Ratio are used to compute score for each opinion bearing words retained after feasibility analysis. For calculating association measures, a set of positive seed-words ($N^{(+)}$) and negative seed-words ($N^{(-)}$) is compiled. The values of positive score ($Score^{(+)}$), negative score ($Score^{(-)}$), and final opinion score ($OpnScore$) of every opinion word are calculated using equations 2, 3, and 4, respectively.

$$Score^{(+)}(w_i) = \sum_{j=1}^{|N^{(+)}|} AssociationFunction(w_i, SeedPos_j) \quad (2)$$

$$Score^{(-)}(w_i) = \sum_{j=1}^{|N^{(-)}|} AssociationFunction(w_i, SeedNeg_j) \quad (3)$$

$$OpnScore(w_i) = Score^{(+)}(w_i) - Score^{(-)}(w_i) \quad (4)$$

If the $OpnScore(w_i)$ value is greater than 0, it is an indication that the word w_i has higher association with positive seed set. Similarly, a negative score refers higher association of the opinion bearing word, w_i with the negative seed set. An opinion score value equal to zero is an indication that the target word, w_i , is equally associated with both positive and negative seed sets, and its polarity behavior is *neutral*.

E. Sentiment Determination using Machine Learning Techniques

As mentioned in the previous section, we have identified a rich set of statistical features including *Pointwise Mutual Information (PMI)*, *Mutual Information (MI)*, *Chi-square*, and *Log Likelihood Ratio (LLR)* to develop an effective word-level sentiment classification system using supervised machine learning approach to determine the sentiment of opinions as *positive*, *negative*, or *neutral*. In addition, we have also considered *negation*, *modifier*, and *tf-idf* features for classification purpose. The proposed method works in two phases – *model learning* and *classification*. The first phase, also called training phase, uses the feature vectors of training dataset to learn a classification model, whereas the classification phase is used to identify semantic orientation of words in new dataset. We have performed experiments with various classifiers for model learning and classification, but finally settled on *Decision Tree (J48)* and *Bagging* algorithms implemented in WEKA due to their best performance. Once the semantic orientation of individual

words is learned, the semantic orientation can be determined at higher levels of abstraction.

IV. EXPERIMENTAL SETUP AND RESULTS

In this section, we present the experimental setup and results of the proposed feature-based opinion mining and sentiment analysis system. The dataset used in our experiment consists of 500 review documents on different models of cell phone crawled from www.amazon.com. Information component extraction mechanism is implemented using rule-based method and anaphora resolution using backtracking technique for identification of candidate $\langle f, m, o \rangle$ pairs. An opinion score generator is implemented in Java to compute the opinion score of the opinionated words retained after *feasibility analysis*. Thereafter, each opinion word is characterized using the features described in section IV. Table 2 presents statistical feature values for a partial list of opinionated words.

It can be observed from Table 2 that majority of the opinion scores obtained using *LLR* is found negative. Thus, it is discarded from further analysis. Thereafter, a Java-based feature vector generator is implemented to generate feature values for each opinion bearing word, and a ternary classification model is learned to classify polarity of a word as *positive*, *negative*, or *neutral*. We have used standard information retrieval performance measures *precision*, *recall*, and *f-score* to analyze the balance between correctness and coverage of the proposed system.

Table 2: Statistical feature values for a partial list of opinionated words

Opinionated word	Statistical feature values			
	PMI	MI	Chi-Square	LLR
Slow	-0.69	-66.15	-369.49	-6.53
Horrible	-0.93	-34.73	-240.89	-8247.24
Bittersweet	0.00	0.00	0.00	0.00
Fantastic	1.46	98.69	607.60	-24541.60
Wonderful	2.04	75.83	419.22	-10558.84

A Evaluating Rule-Based Feature-Opinion Pair Extraction

To the best of our knowledge, no benchmark dataset is available in which features and opinions are marked for electronic products. Therefore, manual evaluation is performed to monitor the overall performance of the proposed system. From the corpus of 500 review documents a total of 75 documents (Digital Camera: 15, iPod: 15, Laptop: 15, and Cell Phone: 30) are randomly selected for testing purpose that contain 1095 sentences. Our rule-based method is applied to extract *feature-opinion* pairs. Initially, the total count obtained for *true positive (TP)*, *false positive (FP)*, and *false negative (FN)* are 396, 567, and 291 respectively. We have observed that, direct and strong relationship between words causes extraction of nouns (or verbs) and adjectives that are not relevant *feature-opinion* pairs. As a result, count of *FP* increases which has an adverse effect on the *precision* value. To overcome this

problem, a Java-based *feasibility analyzer* is implemented to remove noisy *feature-opinion* pairs. After removal of the noisy feature-opinion pairs, the total count of FP reduces to 128. In parallel, we collected all the feature and opinion pairs manually from the test documents. Thereafter, comparing the two sets of pairs *TP*, *FP* and *FN* are calculated. Macro-averaged performance is obtained to present a synthetic measure of performance by simply averaging the result. Table 3 summarizes the performance measure value for our method in the form of a misclassification matrix.

Table 3: Performance evaluation of *feature-opinion* pair extraction using rule-based method

Product	TP	FP	FN	Precision	Recall	F-Score
Camera	70	24	41	0.744	0.630	0.682
iPod	104	23	68	0.818	0.604	0.695
Laptop	27	24	36	0.529	0.428	0.473
Cell Phone	195	57	146	0.773	0.571	0.657
Macro-Average	396	128	291	0.755	0.576	0.654

B. Evaluating Feature-Opinion Binding through Anaphora Resolution

In testing dataset, a total of 891 anaphoric pronouns are present, in which 55 pronouns correctly refer to product features, and 17 pronouns are used as pleonastic occurrence of "it." Our proposed technique for anaphora resolution using backtracking results in an increase of 7.09% in the *recall*. But, at the same time, it results in 8.21% decrease in *precision* value. On analysis, we observed that when context of a pronoun to be resolved exceeds more than two or three sentences, the occurrence of noisy anaphora-antecedent pairs increases. Table 4 summarizes the performance measures of the proposed feature-opinion pairs extraction process through anaphora resolution.

Table 4: Performance evaluation of *feature-opinion* pair extraction after anaphora resolution

Product	TP	FP	FN	Precision	Recall	F-Score
Camera	76	32	35	0.703	0.684	0.694
iPod	111	42	61	0.725	0.645	0.683
Laptop	31	27	32	0.534	0.492	0.512
Cell Phone	209	88	133	0.703	0.611	0.654
Macro-Average	427	189	261	0.693	0.620	0.654

A. Evaluating Sentiment Analysis Process

A Java-based feature vector generator is implemented to generate numeric values of the features for each opinion bearing word. Table 3 shows the information gain ranking of attributes on the basis of WEKA's attribute evaluator, in which the *Chi-square* feature seems to be most discriminative. We have experimented with some prominent classifiers that are best suited for the classification task, but settled with *Bagging* and *Decision Tree (J48)* algorithm due to their best performance. Table 4 shows comparison of four different classifiers in terms of information retrieval metrics on training and testing datasets.

Table 3: Information gain ranking of features

Features	Information Gain
Chi-square	0.3609
MI	0.3542
PMI	0.1680
Negation	0.0657
TF-IDF	0.0236
Modifier	0.0103

Table 4: Classifiers' performance during sentiment analysis

Classifier	Weighted Average Result (over training dataset)			Weighted Average Result (over testing dataset)		
	Precision	Recall	F-score	Precision	Recall	F-score
NB	0.902	0.703	0.779	0.840	0.617	0.683
J48	0.935	0.948	0.939	0.885	0.872	0.855
MLP	0.919	0.940	0.929	0.821	0.857	0.833
Bagging	0.952	0.951	0.942	0.875	0.865	0.848

VI. CONCLUSION AND FUTURE WORK

In this paper, we have presented a feature-based opinion mining and sentiment analysis system. The novelty of the proposed method lies in its enrich set of statistical and NLP features, and their formulation in a way to produce an effective word-level sentiment classification system. Our future works aim to enhance the *precision* of the proposed feature-opinion pairs identification process through anaphora resolution. Devising techniques to deal with noisy and ambiguous anaphora-antecedent binding also seem to be a promising future direction of work.

REFERENCES

- [1] B. Li, L. Zhou, S. Feng, and K. Wong, "A unified graph model for sentence-based opinion retrieval", Proc. 48th Annual Meeting of the Association for Computational Linguistics, Uppsala, Sweden, 2010, pp. 1367–1375.
- [2] B. Pang, and L. Lee, "Opinion Mining and Sentiment Analysis", Foundation and Trends in Information Retrieval, vol. 2, Nos. 1–2, 2008, pp. 1–135.
- [3] G. Qiu, B. Liu, and C. Chen, "Opinion word expansion and target extraction through double propagation", Association for Computational Linguistics 37, 2010, pp.9–27.
- [4] K. Nozomi, R. Lida, K.Inui, Y. Matsumato, "Opinion Extraction Using a Learning-Based Anaphora Resolution Technique", Proc. Second International Joint Conference on Natural Language Processing, 2005, pp. 173-178.
- [5] N. Jakob, I. Gurevych, "Using Anaphora Resolution to Improve Opinion Target Identification in Movie Reviews", Proc. ACL 2010 Conference Short Papers, Stroudsburg, Uppsala, Sweden, 2010, pp. 263-268.
- [6] Bo. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques", Proc. Conference on Empirical Methods in Natural Language Processing (EMNLP), 2002, pp. 79–86.
- [7] S.Kim and E. Hovy, "Determining the sentiment of opinions", Proc. 20th international conference on Computational Linguistics, COLING '04, ACM, 2004.
- [8] A. Kamal, M. Abulaish and T.Anwar, "Mining Feature-Opinion Pairs and their Reliability Scores from Web Opinion Sources", Proc. 2nd Int'l. Conference on Web Intelligence, Mining and Semantics, 2012.