Comparison Extraction Feature Using Double Propagation and Pointwise Mutual Information to Select a Product

A Rahman
Magister of Information System, Universitas Komputer Indonesia, Indonesia
Alkautsar.rahman.s@gmail.com

Abstract. The use of online media such as E-Commerce can allow each person given the freedom to speak his opinions were against the review of a product bought or owned. A review of a product, generally, not only to the product itself but more to its features. For it is done also analyses to be able to take on the features of the product and said his opinions were using the method of Double Propagation (DP) and Pointwise Mutual Information (PMI). The purpose of this research is looking at the right algorithm to find a relevant feature. The resulting feature is very important in making decisions that affect the company. Comparison of methods of DP and PMI based on the results of high accuracy can be the right method.

1. Introduction
Determine the relevant product features to do the proper analysis in a product review. It is used as a reference for the customer to know the opinions of other people about a product or service that will be used in determining the features of a product. [1.2]

Many studies that describe how to find any features that are relevant to understanding the grammatical arrangement of using lexicon based in methods of DP. [3-5] Search features can be easily done with Word's dependency on a sentence with grammar that is raw in the PMI method. [6-8]

Based on previous research the purpose of this scientific work is determining an appropriate and relevant features in a comment or review the product. Comparison of methods of DP and PMI to view the value of accuracy in producing a feature that exists on a review. Advanced handling features and extraction method against the process of granting the orientation of sentiment at every word opinion attached to the product features. The research of using that contain product review laptop using method of DP and PMI. [9-10]

2. Research and Method

2.1. Double Propagation
With the Find feature uses a dictionary of adjectives, then utilizing the features of yesteryear have been extracted to find other adjectives found in the text of the opinions. The new adjectives are automatically added to the dictionary words. This process continuously until there are no new features and the adjectives that were found.

Double Propagation is the approach that is used to perform the extraction of said opinion nor the words product features (target) by using the word iterative opinion and product features that have been
known and have been extracted (at the previous iteration) through identification of syntactic relations. Identification of relationships at Double Propagation method is divided into 3 relationship, namely:

- The relationship between the word and the word opinion product features (OT-rail)
- The relations between the word opinion itself (OO-rail)
- The relationship between the word product features (TT-rail)

DP has four rules and subtask dependencies that can be seen at [9,10]

```
Input: Opinion Word Dictionary \{O\}, Review Data R
Output: All Possible Features \{F\}, The Expanded Opinion Lexicon \{O-Expanded\}
Function:
1. \{O-Expanded\} = \{O\}
2. \{F\} = \emptyset, \{O\} = \emptyset
3. for each parsed sentence in R
4. if(Extracted features not in \{F\})
5. Extract features \{F\} using R1 and R1 based on opinion words in \{O-Expanded\}
6. endif
7. if(Extracted opinion words not in \{O-Expanded\})
8. Extract new opinion words \{O\} using R4 and R4 based on opinion words in \{O-Expanded\}
9. endif
10. endfor
11. Set \{F\} = \{F\} + \{F\}, \{O-Expanded\} = \{O-Expanded\} + \{O\}
12. for each parsed sentence in R
13. if(Extracted features not in \{F\})
14. Extract features \{F\} using R3 and R3 based on features in \{F\}
15. endif
16. if(Extracted opinion words not in \{O-Expanded\})
17. Extract opinion words \{O\} using R2 and R2 based on features in \{F\}
18. endif
19. end for
20. Set \{F\} = \{F\} + \{F\}, \{O\} = \{O\} + \{O\}
21. Set \{F\} = \{F\} + \{F\}, \{O-Expanded\} = \{O-Expanded\} + \{O\}
22. Repeat 2 till size(\{F\}) = 0, size(\{O\}) = 0
```

**Figure 1.** Algorithm Double Propagation.

2.2. **Pointwise Mutual Information**

Semantic relatedness is existing measurements on text mining to describe how relationships between words. The goal of semantic relatedness measurement is to obtain the value or score which represents how much the linkages between those words. PMI is one of the measurements be statistically for the semantic relatedness has been widely used. PMI is a Variant used in this study to calculate relatedness semantic. The general formula of PMI:

$$PMI(w_1, w_2) = \log_2 \frac{P(w_1, w_2)}{P(w_1)P(w_2)}$$

(1)

3. **Results and Discussion**

3.1. **Rules Based Fitur Extraction**

By using the dependency rule sentence can determine a candidate feature on opinion words, such rules can be seen in Figure 2.
Laptop sales review from data consisting of 250 comments we can see the results of the features that have been obtained on table 1.

**Figure 2. Rules for target word extraction.**

**Table 1. Comparison DP and PMI.**

| Features Extraction | Double Propagation | Pointwise Mutual Information |
|---------------------|--------------------|-----------------------------|
| LCD / Display       | 12                 | 5                           |
| Keyboard            | 10                 | 7                           |
| Track Pad           | 5                  | 5                           |
| RAM                 | 16                 | 13                          |
| Harddisk            | 11                 | 10                          |
| Web Cam             | 3                  | 1                           |
| Optical Drive       | 5                  | 4                           |
| Processor           | 19                 | 17                          |
| Battery             | 7                  | 4                           |
| Speaker             | 18                 | 11                          |
4. Conclusions
From table 1 above can be seen a comparison between the method of Propagation and the Double Pointwise Mutual Information, the result of the extraction of features done better using double propagation method, because this method performs a process of iteration in finding a feature. Pointwise Mutual Information method performs the extraction of features by viewing linkages Word closely related with the word after or before, this has resulted in a lack of accuracy in the eject feature in a product review.

References
[1] Qiu G, Liu B, Bu J and Chen C 2011 Opinion word expansion and target extraction through double propagation Computational linguistics 37(1) 9-27
[2] Recchia G and Jones M N 2009 More data trumps smarter algorithms: Comparing pointwise mutual information with latent semantic analysis Behavior research methods 41(3) pp. 647-656
[3] Zhang L, Liu B, Lim S H and O'Brien-Strain E 2010 Extracting and ranking product features in opinion documents Proceedings of the 23rd international conference on computational linguistics: Posters pp. 1462-1470 Association for Computational Linguistics
[4] Qiu G, Liu B, Bu J and Chen C 2009 Expanding domain sentiment lexicon through double propagation IJCAI 9 pp. 1199-1204
[5] Pang B and Lee L 2008 Opinion mining and sentiment analysis Foundations and Trends® in Information Retrieval 2(1–2) 1-135
[6] Su Q, Xiang K, Wang H, Sun B and Yu S 2006 Using pointwise mutual information to identify implicit features in customer reviews International Conference on Computer Processing of Oriental Languages 8 pp. 22-30
[7] Xu Y, Jones G J, Li J, Wang B and Sun C 2007. A study on mutual information-based feature selection for text categorization Journal of Computational Information Systems 3(3) pp.1007-1012
[8] Schneider K M 2005 Weighted average pointwise mutual information for feature selection in text categorization European Conference on Principles of Data Mining and Knowledge Discovery 8 pp. 252-263
[9] Zhang L and Liu B 2014 Aspect and entity extraction for opinion mining Data mining and knowledge discovery for big data 5 pp. 1-40
[10] Qiu G, Liu B, Bu J and Chen C 2011 Opinion word expansion and target extraction through double propagation Computational linguistics 37(1) pp. 9-27