JURNAL UPI

Journal of Software Engineering, Information and Communication Technology (SEICT)

Journal homepage: https://ejournal.upi.edu/index.php/SEICT

Pointwise Mutual Information for Opinion Mining Feature Level Product Reviews

Indira Syawanodya

Informatics Engineering, Universitas Telkom, Indonesia
Correspondence: E-mail: indirasyawanodya@gmail.com

ABSTRACT

The determination of whether a product holds a positive or negative sentiment can be inferred from reviews provided by previous customers. In recent years, various websites have emerged that offer product reviews, in which the value of a product is evaluated through user-generated ratings and textual comments. However, the abundance of reviews often makes it challenging for prospective customers to interpret the overall sentiment accurately. To address this issue, a classification approach can be employed to determine the polarity of product reviews. Opinion Mining, also known as sentiment analysis, is a field of study that focuses on analyzing individuals' opinions toward entities, individuals, issues, events, topics, and their associated attributes. The implementation of feature extraction prior to the classification process has been shown to significantly enhance the accuracy of sentiment assessment. One effective method for feature extraction is Pointwise Mutual Information (PMI), which leverages search engine statistics to identify meaningful term associations in real time. PMI enables the system to capture semantic relationships between words, thereby improving the reliability of sentiment classification.

ARTICLE INFO

Article History:
Submitted/Received 04 Apr 2025
First Revised 28 Apr 2025
Accepted 10 May 2025
First Available online 01 Jun 2025
Publication Date 01 Jun 2025

Keyword:
Feature extraction;
Opinion mining;
Pointwise mutual information;
Product review;
Review opinion.

1. INTRODUCTION

Today, product reviews are commonly available online, especially on e-commerce platforms. These reviews typically consist of ratings or written comments provided by customers. Such feedback benefits not only producers but also other customers (Mudambi et al, 2010). Producers can analyze the strengths and weaknesses of the products they offer, while customers can learn about the pros and cons before making a purchase. However, due to the large number of reviews, it can be difficult for people to draw meaningful conclusions (Saphira O, et al, 2020). Therefore, a system is needed to determine whether a product review expresses a positive or negative sentiment. The classification process can assist in identifying the overall opinion conveyed in a review (Liu B, 2012).

Opinion Mining is the study that focuses on analyzing people's opinions toward entities, individuals, issues, events, topics, and their attributes (Liu B, 2012). One method used to determine the polarity of opinions is the Semantic Orientation Label. Performing feature extraction before classification can significantly improve the accuracy of sentiment analysis. Pointwise Mutual Information (PMI) is commonly used in feature extraction, as it calculates the co-occurrence frequency of terms by retrieving real-time data from search engines (Manning C.D. et al, 2008).

This work is closely related to the study conducted by Ana-Maria Popescu (2007), which proposed ten extraction rules to identify opinion phrases. In the current study, only four of the original ten rules are adopted to simplify the extraction process. In the earlier work, feature extraction was performed using MINIPAR, a syntactic parser that analyzes the grammatical structure of reviews.

2. METHODS

2.1. Pre-processing

Pre-processing is conducted to identify candidate features. In this study, the Stanford Parser is employed to extract noun phrases from each product review (Manning et al, 2014). The Stanford Parser analyzes the grammatical structure of a sentence and is capable of processing simple textual inputs to produce various outputs, including part-of-speech tags, syntactic parse trees, and typed dependencies (de Marneffe et al, 2014).

To refine the candidate features, several pre-processing techniques are applied, namely stopword removal, lemmatization, and part-of-speech (POS) tagging. Stopword removal is used to eliminate non-informative words such as "I," "you," "very," and others. Lemmatization is a natural language processing technique used to retrieve the base or root form of a word by utilizing a combination of corpus-based and morphological analysis (Balakrishman et al, 2014). For example, the word broken will be converted to break through the lemmatization process. POS Tagging assigns a syntactic category or tag to each word in a sentence (Santorini et al, 1991). For instance, the sentence "the phone has good sound volume" would be tagged as: the_DT phone_NN has_VBZ good_JJ sound_NN volume_NN.

2.2 Feature Extraction

Pointwise Mutual Information (PMI) is used to identify candidate features by extracting noun phrases from the reviews. The noun phrases obtained from the pre-processing stage are then evaluated using the PMI algorithm, as defined by the following equation:

DOI: https://doi.org/10.17509/seict.v6i1.89411
p- ISSN 2774-1656 e- ISSN 2774-1699

$$score(choice_i) = \frac{hits(problem\ AND\ choice_i)}{hits(choice_i)}$$

In this formula, $hits(problem\ AND\ choice_1)$ represents the number of search engine results returned for the co-occurrence of the problem and the choice. Here, problem functions as a discriminator for $choice_1$, and choice refers to a phrase (Turney et al, 2001). The value obtained from the equation, referred to as the score, is compared with a predetermined threshold. If the score exceeds the threshold, the candidate feature is considered a valid feature.

The resulting score from the equation is then compared against a predetermined threshold. If the score exceeds the threshold, the candidate feature is considered a valid (real) feature.

2.3 Find Opinion and Their Polarity

The sentence, which is include real feature, is looked for phrase that have a potential opinion by OPINE extraction rule. OPINE is an unsupervised information extraction system which mines review to build a model from product review. OPINE has an extraction rule to indicate potential opinion phrase. The extraction rule is showed below.

Table 1. OPINE Extraction Rule

Rule Extraction	Example
$if\ \exists(M, NP = f) \longrightarrow po = M$	(expensive) scanner
$if\ \exists(S = f, P, O) \longrightarrow po = O$	lamp has (problems)
$if\ \exists(S, P, O = f) \longrightarrow po = P$	I (hate) this scanner
$if\ \exists(S = f, P, O) \longrightarrow po = P$	program (crashed)

To find polarity from each opinion, SO label can determine a semantic orientation by search engine in real,time. SO label is an unsupervised technic. To find polarity, this algorithm gets score from word w by SO Label equation (Turney et al, 2001).

$$SO(phrase) = log_2\left(\frac{hits(a)hits(b)}{hits(c)hits(d)}\right)$$

Where :

a : phrase NEAR “excellent”
b : “poor”
c : phrase NEAR “porr”
d : “excellent”

The reference word excellent and poor is a limit of score. Generally, rating 1 is defined by ‘poor’ and rating 5 is defined by ‘excellent’. Value SO positive if phrase score is bigger than “excellent”, and negative if phrase score is lower than “poor” (Turney et al, 2002).

DOI: https://doi.org/10.17509/seict.v6i1.89411
p- ISSN 2774-1656 e- ISSN 2774-1699

2.4. Post-processing

Postprocessing is a way to identify patterns from application (Liu B, 2007). With postprocessing, extracted features will be filtered by comparing them with the dataset corpus.

3. RESULTS AND DISCUSSION

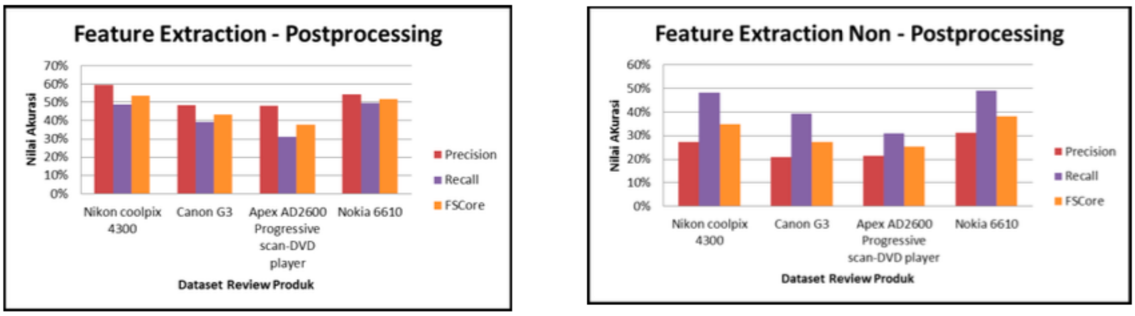
This paper use dataset from Hu Liu which has been used in previous research (Hu M et al, 2006). Precision and recall use to measure the accuracy of data (Hu M et al, 2006).

$$Precision = \frac{\sum EC_i}{\sum E_i}$$

$$Recall = \frac{\sum EC_i}{\sum C_i}$$

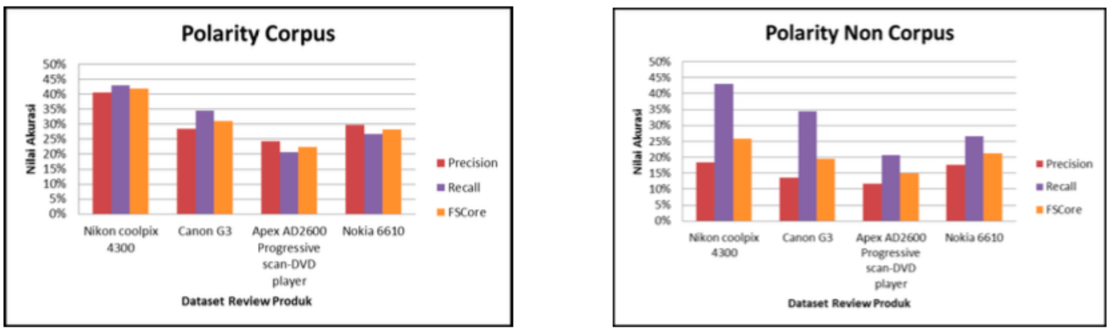
Figure 1 shows the results of the initial experiment comparing feature extraction with and without post-processing. From the figure, it can be observed that post-processing has a significant impact on the results. The accuracy increases by approximately 15%.

Post-processing not only affects feature extraction but also influences the polarity score, as the system identifies opinion phrases based solely on the predefined corpus. This effect is illustrated in Figure 2.



(a) (b)

Figure 1. Feature Extraction (a) Non Post Processing, (b) Post Processing



(a) (b)

Figure 2. Polarity (a) Non Post Processing, (b) Post Processing

In the second experiment, this study analyzes the impact of the threshold value setting. The results show that each dataset has a different optimal threshold value. Figure 3 presents the precision accuracy for four reviews, while Figure 4 illustrates the recall.

DOI: https://doi.org/10.17509/seict.v6i1.89411
p- ISSN 2774-1656 e- ISSN 2774-1699

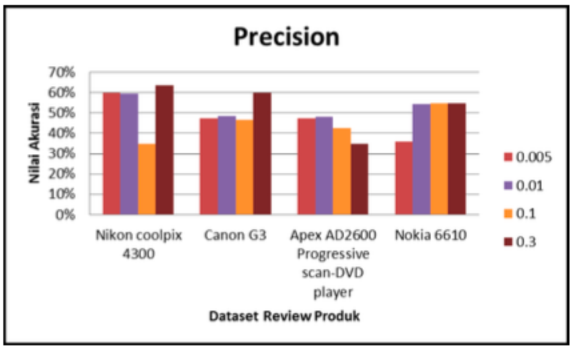


Figure 3. Precission Result

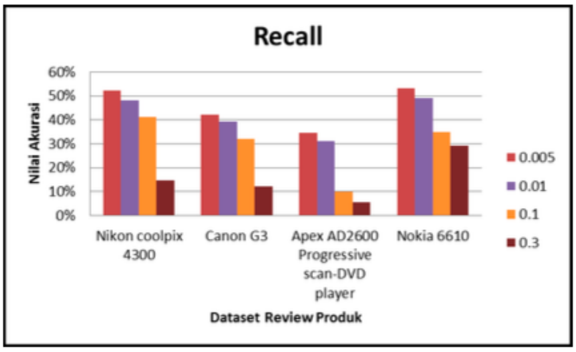


Figure 4. Recall Result

From Figure 3, it can be observed that each dataset has its own optimal threshold value. In this study, four threshold values were compared: 0.005, 0.05, 0.1, and 0.3. However, in terms of recall, smaller threshold values tend to result in higher accuracy. This occurs because a greater number of relevant items are successfully identified when the threshold is lower.

It is important to note that lowering the threshold increases the likelihood of identifying more candidate features, which contributes to higher recall. However, this may come at the cost of reduced precision, as more non-relevant items may also be included. Conversely, higher threshold values may improve precision by filtering out irrelevant items but risk missing some relevant features, thereby reducing recall. This reflects the common trade-off between precision and recall in information retrieval and classification tasks.

4. CONCLUSION

Based on the experimental results, the conclusions of this study are as follows: post-processing affects not only feature extraction but also the overall classification results. It can increase up to 15%. Furthermore, the choice of threshold value also influences the precision and recall accuracy for each dataset.

For future work, researchers may explore methods to identify implicit features from reviews, which are not explicitly mentioned but can be inferred through contextual understanding

5. AUTHORS' NOTE

The authors declare that there is no conflict of interest regarding the publication of this article. Authors confirmed that the paper was free of plagiarism.

6. REFERENCES

Balakrishman, V., and Ye, E. (2014). Stemming and lemmatization: A comparison of retrieval performances.

de Marneffe, C., and Manning, C. D. (2013). Stanford typed dependency manual. Stanford University.

DOI: https://doi.org/10.17509/seict.v6i1.89411
p- ISSN 2774-1656 e- ISSN 2774-1699

Etzioni, O., Cafarella, M., Downey, D., Popescu, A.-M., Shaked, T., Soderland, S., and Weld, D. S. (2005). Unsupervised named-entity extraction from the web: An experimental study. Department of Computer Science and Engineering, University of Washington.

Mudambi, S. M., and Schuff, D. (2010). What makes a helpful online review? A study of customer reviews on Amazon.com. MIS Quarterly, 34(1), 185–200

Hu, M., and Liu, B. (n.d.). Mining and summarizing customer reviews. University of Computer Science.

Hu, M., and Liu, B. (2006). Opinion feature extraction using class sequential rules. In Proceedings of the American Association for Artificial Intelligence (AAAI).

Liu, B. (2007). Web data mining: Exploring hyperlinks, contents, and usage data. Springer.

Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1), 1–167. Morgan & Claypool Publishers.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to information retrieval. Cambridge University Press.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., & McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, 55–60. Association for Computational Linguistics (ACL).

Popescu, A.-M., and Etzioni, O. (2007). Extracting product features and opinions from reviews. In Proceedings of NLP and Text Mining (pp. 9–28).

Santorini, B. (1991). Part-of-speech tagging guidelines for the Penn Treebank Project. University of Pennsylvania.

Shapira, O., and Levy, R. (2020). Massive multi-document summarization of product reviews with weak supervision. arXiv preprint arXiv:2007.11348. Cornell University Library.

Turney, P. D. (2001). Mining the web for synonyms: PMI-IR versus LSA on TOEFL. In Proceedings of the Twelfth European Conference on Machine Learning (ECML-2001) (pp. 491–502).

Turney, P. D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL) (pp. 417–424).

DOI: https://doi.org/10.17509/seict.v6i1.89411
p- ISSN 2774-1656 e- ISSN 2774-1699