International Journal of Computer & Software Engineering

Original Article **Open Access**

Development of a Part-of-Speech Combination Method for Chinese eWOM Analysis

Yuh-Jen Chen[1*], Yuh-Min Chen[2], Sheng-Chieh Kao[2] and Jyun-Han Wu[2]

[1]Department of Accounting and Information Systems, National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan, ROC
[2]Graduate Associate, Institute of Manufacturing Information and Systems, National Cheng Kung University, Tainan, Taiwan, ROC

Abstract

Electronic word-of-mouth (eWOM) has recently become a common channel for spreading product appraisals and an important reference for enterprise internal improvement. Nevertheless, the various virtual communities that have emerged with the popularity of the Internet have also resulted in information overload because of the explosive growth of eWOM. Therefore, how to help enterprises analyze useful decision references from a large amount of eWOM has become a key issue in implementing customer relationship management (CRM).

This study develops a method of Part-of-Speech (POS) combination to support eWOM analysis to effectively help an enterprise rapidly and accurately realize the contents of eWOM appraisals to improve customer relationships. This objective can be achieved by performing the following tasks: (i) designing a process for POS-combination supported eWOM analysis, (ii) developing techniques related to POS-combination supported eWOM analysis, and (iii) implementing a mechanism for POS-combination supported eWOM analysis.

Introduction

Numerous enterprises have recently promoted customer relationship management (CRM) to maintain favorable interaction with customers and enhance their reputation and profits. Customers' word-of-mouth (WOM) has been integral to CRM. The popularity and development of social media has changed past consumer WOM and behaviors. Customer participation in WOM interaction through the Internet has increased the importance of electronic word-of-mouth (eWOM) [2,17]. However, the explosive growth of the amount of eWOM information has caused a dilemma of information overload [4,12]. Consequently, effectively analyzing valuable decision references from large amounts of eWOM information has become a key issue for enterprises in implementing CRM.

Previously, single polarity terms (ex. "delicious") [8,11,15], continuous terms N-gram (ex. "free upgrade") [10,19] or negatives [14] (ex. "not tasty") were used as judgment features in analyzing eWOM appraisals. However, the interaction between terms and syntax affects semantic polarity. In linguistics, semantic knowledge was observed, collected, and organized from the two aspects of paradigmatic and syntagmatic relation [1]. The former described the taxonomy of words, like the positive words "convenient", "delicious", "healthy", and "favorable". Meanwhile, the latter emphasized the combination of terms into larger linguistic units, such as the relations among "golden", "crispy", and "skin" in the sentence "golden and crispy skin". Syntax decided the permutation combination of words in a sentence [9]. Additionally, eWOM analyses previously ignored the descriptive topic in an article [10,15]. Such analytical results can not correspond to the target enterprise and products. Besides, vocabulary and syntax of popular colloquial languages often appeared on the descriptions in eWOM articles, which forced constant update of inner judgment appraisal in eWOM analyses.

This study develops a method of Part-of-Speech (POS) combination to support eWOM analysis, in consideration of POS-combination feature, topic identification, and constant update of appraisal judgment, to effectively help an enterprise rapidly and accurately realize the contents of eWOM appraisals to improve customer relationships.

This objective can be achieved by performing the following tasks: (i) designing a process for POS-combination supported eWOM analysis, (ii) developing techniques related to POS-combination supported eWOM analysis, and (iii) implementing a mechanism for POS-combination supported eWOM analysis. Developing techniques associated with POS-combination supported eWOM analysis involves POS combination rule generation, topic term extraction, and eWOM identification.

Design of POS-Combination Supported eWOM Analysis Process

This section designs the process of POS-combination supported Chinese eWOM analysis. As shown in Figure 1, the process is composed of POS-combination rule generation, topic term extraction, and eWOM identification, as presented below.

POS-Combination rule generation and topic term extraction

POS-combination rule generation and topic term extraction involves four phases of POS-combination classification, polarity POS-combination identification, POS-combination rule generation, and topic term library establishment, as described below.

POS-combination classification: The POS-combination in Chinese eWOM articles is innovative and changeable, and thus the major POS-combination type is not easily determined. This study first determines the POS-combination type which might contain polarity.

*Corresponding Author: Prof. Yuh-Jen Chen, Department of Accounting and Information Systems, National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan, ROC, Tel: +886-7-6011000 ext. 34316, Fax: +886-7-6011158; E-mail: yjchen@nkust.edu.tw

Non-polarity sentences are then filtered through known polarity terms, and lengthy sentences are ignored. However, the structure of Chinese sentences is often scattered. Thus, similar POS-combinations need to be clustered to identify a representative POS-combination type.

Polarity POS-combination identification: Since polarity sentences in Chinese eWOM articles obviously express customer perceptions for the target enterprise or products. Hence, determining POS-combinations with polarity can facilitate the analysis of Chinese eWOM. In the phase, the polarity terms are first sifted from the example sentences with the same POS-combination, after which the polarity POS-combination is analyzed to acquire the polarity POS-combinations with high-frequency of occurrence.

POS-combination rule generation: Same polarity keywords can generate different meanings in different POS-combinations [14], such as "比較(ADV)新鮮(Vi)好吃(Vi)" and "不能(N)說(Vt)是(Vt)不好(ADV)吃(Vi)". Therefore, polarity rules from different polarity POS-combinations can enhance the accuracy of eWOM analysis. During this phase, sentences with the same POS-combination are first sought from the library of articles with known polarity sentences. These sentences with the same POS-combination are then matched with the library of polarity term to determine the polarity of polarity terms. Moreover, the polarity relationship between polarity terms and sentences is judged to derive rules of the polarity POS-combination.

Topic term library establishment: Topics in Chinese eWOM articles frequently change and are not completely described for a single subject. Thus, it is necessary to confirm whether the described subject in the sentence is the target product of an enterprise. In this study, the library of topic term is established (such as the relevant topic terms in the pastry industry, namely "酥鬆(crispy)", "美味(delicious)", and "伴手(local specialty gift)") and used to identify the described subject in the sentence, and ultimately confirm the described subject in Chinese eWOM.

eWOM Identification

Phases in eWOM identification include topic confirmation for sentences and polarity analysis. Each phase is explained as follows.

Topic confirmation for sentences: The phase is to confirm the topic of sentences using the library of topic term to judge whether the unknown topic sentence is the target subject. Previous studies first obtained relevant topic terms from possible topic types [14], and the relevant topic terms from different topics were then used to judge the different topics of sentences. This method can not search for all unknown future topics, and thus is modified and used in this study to acquire the topic term of the target topic as a reference for judging the topic.

Polarity analysis: In the polarity analysis, the library of POS-combination rule and the library p of polarity term are utilized to analyze the WOM appraisal of sentences. First, POS-combinations of sentences conforming to the topic are classified based on the library of POS-combination rule. The POS-combination type classified into the library of POS-combination rule confirms whether the type has polarity by the library of POS-combination rule. Furthermore, the term polarity of polarity keywords is retrieved from the library of polarity term, and the retrieved polarity of polarity keywords is combined with POS-combination rules to determine the polarity of the sentence. If the polarity term does not appear in the library of polarity term, the term polarity is analyzed by the knowledge constructor to acquire the term polarity of the keyword. If the sentence can not be classified into known POS-combinations, antonym logic rules are applied to analyze the eWOM appraisal.
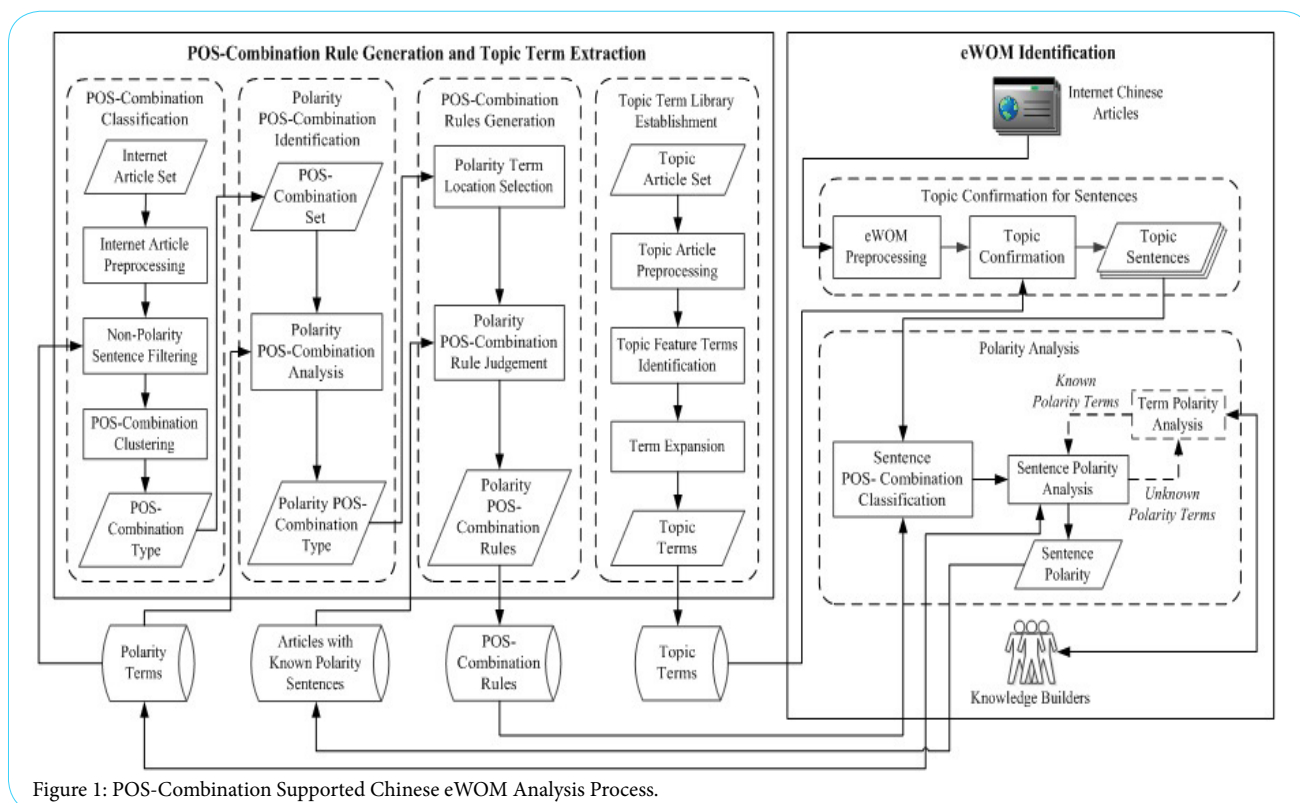

Figure 1: POS-Combination Supported Chinese eWOM Analysis Process.

Development of POS-Combination Supported Chinese eWOM Analysis Techniques

Based on the process of POS-combination supported Chinese eWOM analysis designed in Section 2, this section develops techniques for POS-combination supported Chinese eWOM analysis to assist enterprises rapidly and accurately understand current eWOM appraisals. These techniques include POS-combination rule generation and topic term extraction and eWOM identification, all of which are discussed below.

Pos-combination rule generation and topic term extraction

POS-combination rule generation and topic term extraction involves POS-combination classification, polarity POS-combination identification, POS-combination rule generation, and topic term library establishment, as detailed below.

POS-combination classification

POS-combination classification comprises the three steps of Internet article preprocessing, non-polarity sentence filtering, and POS-combination clustering. Each step is explained as follows.

Internet Article Preprocessing

As illustrated in Figure 2, the CKIP system [3,13] is first used for sentence and word segmentation and POS tagging to obtain the POS structure of these sentences. The obtained POS structures are then viewed as POS-combination features.

Sentence segmentation and POS tagging: Chinese Knowledge and Information Processing (CKIP) with the functions of sentence and word segmentation and POS Tagging is adopted to analyze the POS structure of Chinese sentences, such as "這樣(Vi)的(T)美食(N)能(ADV)錯過(Vt)嗎(T)".

Proper noun tagging: For the results of POS tagging, some proper nouns did not appear in the word library of the CKIP system, which results in wrong POS tagging. A list of proper noun tag thus is artificially established to revise the proper noun tag of the enterprise and products. For instance, the Chinese sentence "千萬(Neu)別(D)錯過(VJ)這(Nep)個(Nf)芋頭(Na)酥(VH)啦(T)" has wrong POS tags, and the revised POS tag is "芋頭酥(Na)".

Antonym and confused term tagging: In the CKIP system, antonym terms and confused terms are not obviously tagged, which easily cause errors in polarity analysis. Hence, a list of antonym term and confused term tag is artificially established to revise the POS tag. For example, "不(ADV)" is revised as "不(ADV1)" and "不(ADV1)小心(Vt)" is revised as "不小心(Vt)".

Non-Polarity Sentence Filtering

Since this study merely clusters POS-combination sentences with polarity, the library of polarity term is used to filter non-polarity sentences and keep polarity sentences with polarity terms. Figure 3 presents the process of non-polarity sentence filtering.

Polarity sentence filtering: In this step, prepossessed sentences are compared with polarity terms in the polarity term library to keep the prepossessed sentences that include the polarity terms and the POS-combination features. For example, "好" in the prepossessed sentence "服務(Nv)真(ADV)好(Vi)" is the polarity term in the library of polarity term. The POS-combination "(Nv)(ADV)(Vi)" is kept.

POS-combination count: The number of generated POS-combinations is counted, e.g. the POS-combination "(Nv)(ADV)(Vi)" appears seven times.

POS-combination clustering

Clustering is utilized to integrate types of POS-combination. POS-combinations can be clustered through their permutations. Similar
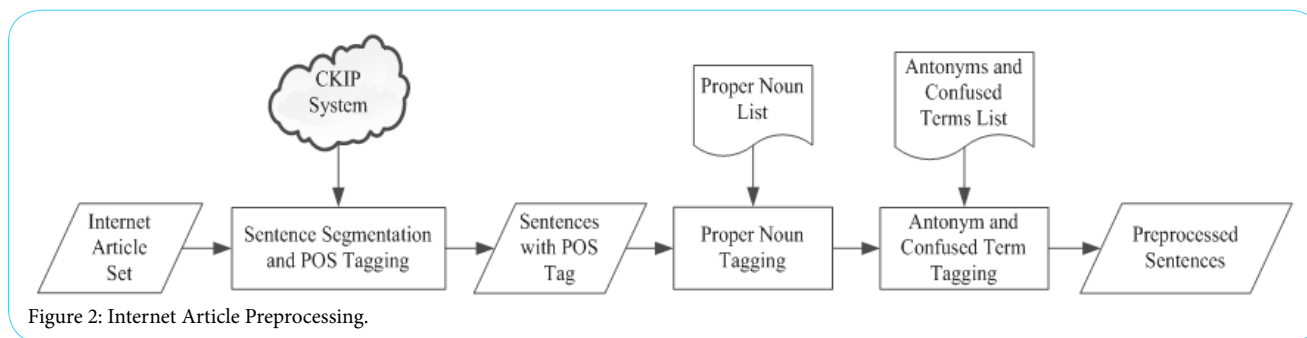
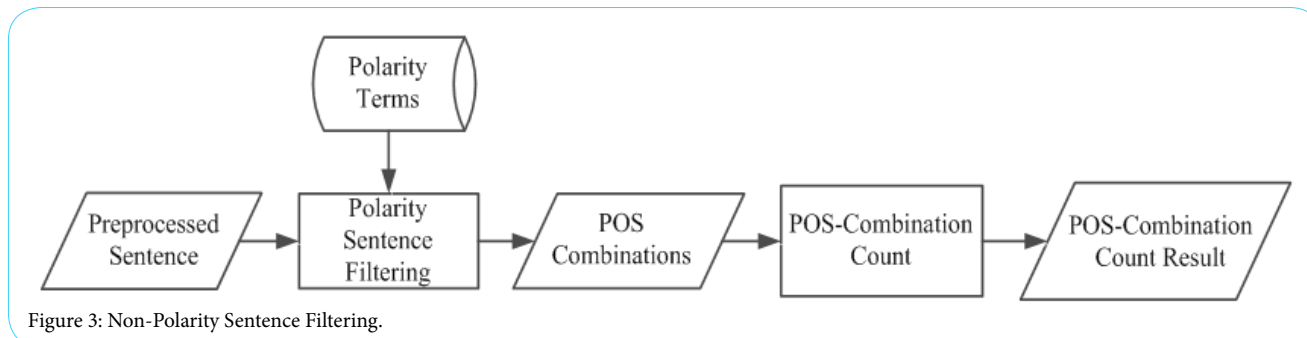

Figure 2: Internet Article Preprocessing.

Figure 3: Non-Polarity Sentence Filtering.

POS-combinations are regarded as the same type, and the semantic rules of such a type are determined to support the analysis of Chinese eWOM. The steps involves POS-combination clustering and POS-combination structure clustering, as illustrated in Figure 4.

POS-combination clustering: FDOD (Function of Degree Of Disagreement) [7,20] is used and modified to cluster POS-combinations. The similarity between POS-combinations is calculated using the revised formula in Equation (1).

$$FDOD(S_1, S_2) = \sum_{i=1}^{n} cis(S_1, H_n)^* cis(S_2, H_n) \qquad (1)$$

where $FDOD(S_1, S_2)$ denotes the difference between POS-combinations

S_1 and S_2 calculated using $FODD$;
S is the set of POS-combinations, expressed as $S \in \{S_1, S_2, S_3, ..., S_x\}$;
L is window size of per unit, which is normally set to 2;
H is the set of sequences from cutting two sequences with length L;
H_n is the content in the n-th unit of set H;
$cis(S, H_n)$ denotes the number of POS being continuously same in matching S with H_n;

In Equation (8), the similarity between longer sequences is easily enhanced while the equation is applied to calculate the similarity between shorter POS-combinations. This study designs a weighted method that divides the results of in Equation (8) with the possible combinations that are generated by pair of sentences with POS. Each pair of sentences is expected to have a close POS quantity and a different penalty value of POS quantity. The formula is written as Equation (2).

$$FDODPatch(S_1, S_2)$$
$$= FDOD(S_1, S_2)^* C_1 / (len(S_1)^* len(S_2)) - |(len(S_1) - len(S_2))|^* C_2 \quad (2)$$

where $FDODPatch(S_1, S_2)$ represents the standardization of the results that are calculated by $FDOD(S_1, S_2)$;

C_1 and C_2 are fixed constants;
$len(S)$ denotes the length of POS quantity in POS combination S;

Subsequently, the results that are calculated using Equation (2) undergo Density-based Spatial Clustering of Applications with Noise (DBSCAN) [5,6] to achieve POS-combination clustering.

POS-combination structure clustering: Since the matching of POS-combination structure can be regarded as a problem of sequence alignment, dynamic programming [16,18] is thus applied to cluster POS-combination structures based on the results of POS-combination clustering. Sequence alignment determines the alignment positions between different POS combinations so that different POS-combinations have a uniform position of polarity term. The calculation is detailed as follows.

Assume that each POS tag in POS-combination a is expressed as $(a_1, a_2, ..., a_i)$ and each POS tag in POS-combination b is expressed as $(b_1, b_2, ..., b_i)$, the score of each sub-question is recorded as matrix A. The element $A(i,j)$ is calculated using Equation (3).

$$A(0,0) = 0$$
$$A(i,0) = -i$$
$$A(0,j) = -j \qquad (3)$$
$$A(i,j) = \begin{cases} \max \begin{cases} A(i-1, j-1) - 1 \\ A(i-1, j) - 1 \\ A(i, j-1) - 1 \end{cases} & f\ a_i \neq b_j \\ A(i-1, j-1) + 2 & f\ a_i = b_j \end{cases}$$

Finally, POS-combination types are clustered using the DBSCAN clustering algorithm according to the similarity score between two sequences.

Polarity POS-combination identification

Based on the clustered POS-combination types, polarity POS-combination is analyzed. The analysis utilizes the library of articles with known polarity sentences (including known polarity POS-combination) to identify the polarity POS-combinations that must be judged, as shown in Figure 5.

POS-combination example sentence obtainment: POS-combinations in a single POS-combination type are compared with sentences in Internet articles to extract sentences with the same POS-combination.

Polarity N-gram tagging: An N-gram POS-combination is the combination of continuous N POS appearing in a sentence, such as two of the continuous POS-combinations in Table 1. Since a specific POS-combination in an N-gram may have polarity, an N-gram POS-combination is used as the feature tag of polarity POS-combination.
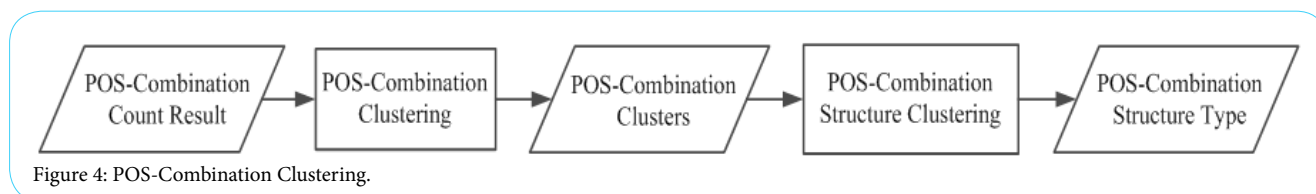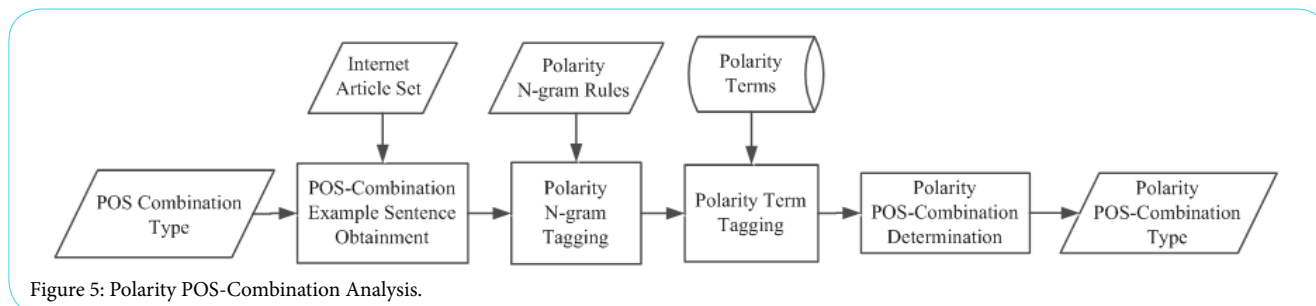

Figure 4: POS-Combination Clustering.

Figure 5: Polarity POS-Combination Analysis.

Item	First Term	Second Term
1	Adverb	Verb
2	Adjective	Noun
3	Adverb	Adjective
4	Noun	Adjective

Table 1: Polarity 2-gram POS-combination.

Polarity term tagging: The feature of the polarity term can be used to judge whether the POS-combination is a polarity POS-combination. The feature of the polarity term is firstly acquired from the library of polarity term. Then, the POS-combination example sentences that may have the polarity are tagged if the POS-combination example sentences contain the feature of the polarity term.

Polarity POS-combination determination: A POS-combination type with/without polarity can be determined based on the proportion of polarity that tagged in POS combination example sentences. In this study, the proportional threshold is set to 80% for the more accurate identification of polarity POS-combination types.

POS-Combination rules generation

This subsection analyzes the determined polarity POS-combinations to generate polarity POS-combination rules. The POS-combination rule generation includes polarity term location selection and polarity POS-combination rule judgment, as detailed below.

Polarity term location selection

The selection of polarity term location facilitates the accuracy of analyzing Chinese eWOM. Figure 6 depicts the detailed steps in selection of polarity term location.

POS-combination sentence obtainment: POS-combinations in a single polarity POS-combination type are matched with sentences in Internet articles to extract sentences with the same POS-combination.

Polarity term location tagging: The feature of the polarity term is firstly acquired from the library of polarity term, and its location in the POS-combination is then tagged for polarity POS-combination example sentences.

Polarity term location judgment: The probability of each location in polarity POS-combination example sentences that are tagged with polarity is calculated. When the probability of the location with the feature of the polarity term exceeds the threshold T_w, the location is regarded as the location of an emotional word.

Polarity POS-combination rule judgment

The judgment of the polarity POS-combination rule mainly analyzes whether the sentence meaning is changed while the polarity term appears in polarity POS-combination. For example, while the polarity term "好吃" appears in a POS-combination example sentence "不能(N)說(Vt)是(Vt)不(ADV)好吃(Vi)", the sentence meaning tends to be positive. Figure 7 presents the steps in judging the polarity POS-combination rule.

Sentence polarity tagging: According to the POS-combination, location of the polarity term, and sentences in the library of articles with known polarity sentences, POS-combinations are matched to extract those with the same POS-combination.

Term polarity tagging: The example sentences and the location of the polarity term that were acquired in tagging of sentence polarity are used to determine the polarity terms. The polarity terms thus obtained are then matched with those in the library of polarity term, and the polarities of those term are recorded.

Polarity POS-combination rule judgment: The meaning of the example sentence is matched with the term polarity to record whether the meaning of the example sentence and the term polarity are the same or opposite. In the process, if multi-polarity terms appear, they are combined on the basis that a double negative is a positive. The
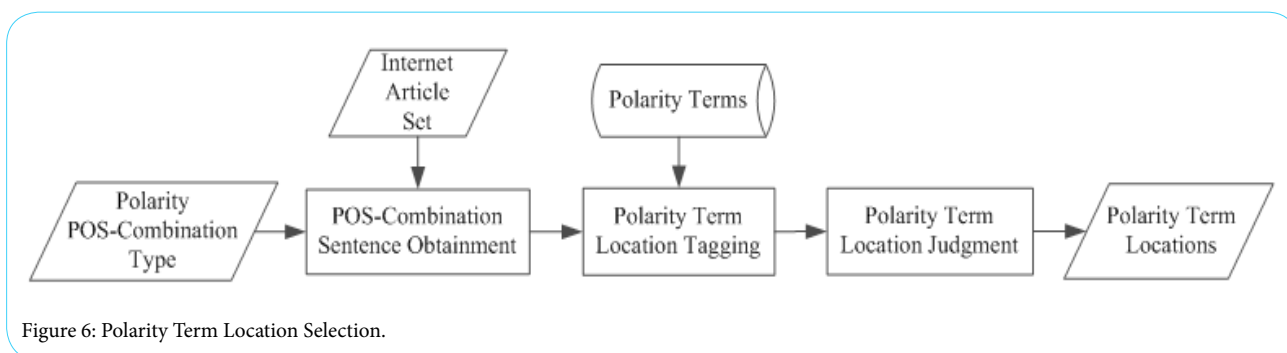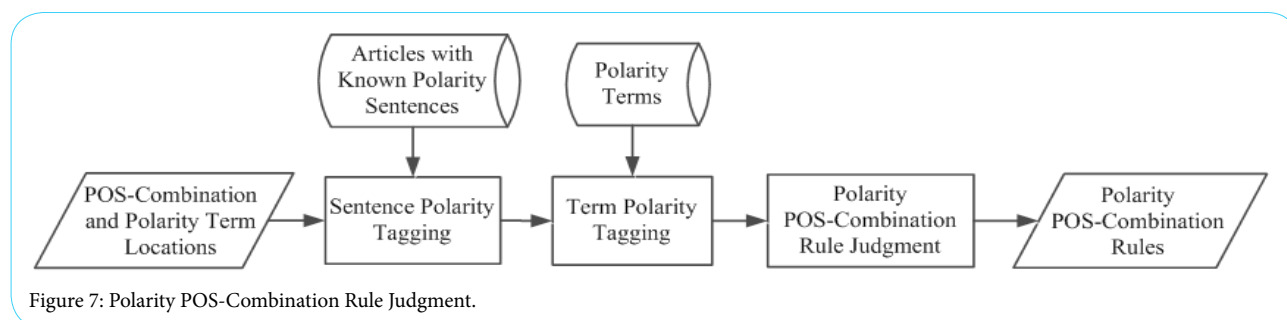

Figure 6: Polarity Term Location Selection.

Figure 7: Polarity POS-Combination Rule Judgment.

example sentences in the same polarity POS-combination type are processed using probability and statistics. When the same or the opposite probability exceeds the 70% threshold, the POS-combination rule for the POS-combination type is judged; otherwise, the POS-combination type is ignored.

Topic term library establishment

The establishment of topic term library includes topic article preprocessing, topic feature term identification, and term expansion, as explained below.

Topic article preprocessing

Articles in the set of topic article, regardless of whether they conform to the topic, are preprocessed to acquire meaningful term units. The process of topic article prepossessing is the same as that of Internet article preprocessing described before (Figure 2).

Topic feature term identification

The identification of the topic feature term involves the selection of more important terms related to topics based on the prepossessed terms obtained from the topic article preprocessing, as shown in Figure 8.

In Figure 8, topic term selection utilizes term frequency-inverse document frequency (TF-IDF) to examine the importance of topic terms in the topic article. Equation (4) presents the formula for TF-IDF.

$$W_{i,j} = tf_{i,j} * log\left(\frac{N}{df_i}\right) \qquad (4)$$

where $W_{i,j}$ denotes the importance of vocabulary i appearing in topic article j;

TF denotes the frequency of vocabulary appearing in the topic article, with higher frequency indicating higher importance of the vocabulary in the topic article;

$IDF = log\left(\frac{N}{df_i}\right)$ represents the frequency of vocabulary appearing in all topic articles. Lower frequency indicates better representativeness of the vocabulary, and thus that the article is easy to differentiate;

i is a vocabulary;
j is the article appearing vocabulary i;
N is the total of all topic article types;
Df is the number of articles in which vocabulary i appears;

According to the importance of topic terms calculated by TF-IDF, expletives that are the same as stop words and meaningless terms are removed. The remaining terms, with importance value W higher than the threshold, are then selected as the topic feature terms.

Term expansion

Term expansion is to expand the relevant terms for the topic feature terms selected from step (2) to derive keywords that do not appear in the article but are related to the topic. The detailed steps are shown in (Figure 9).

Context clue expansion: Google N-gram [21] is used to search for the terms frequently appearing in front of or behind the topic term. For instance, the topic term "mobile" is given for Google 5-gram, and the search results may be "the mobile screen is too small". Subsequently, the number of occurrences of the terms is counted, and terms that are the same as stop words are removed. Finally, a threshold is set to keep the terms whose occurrence number exceeds the threshold.
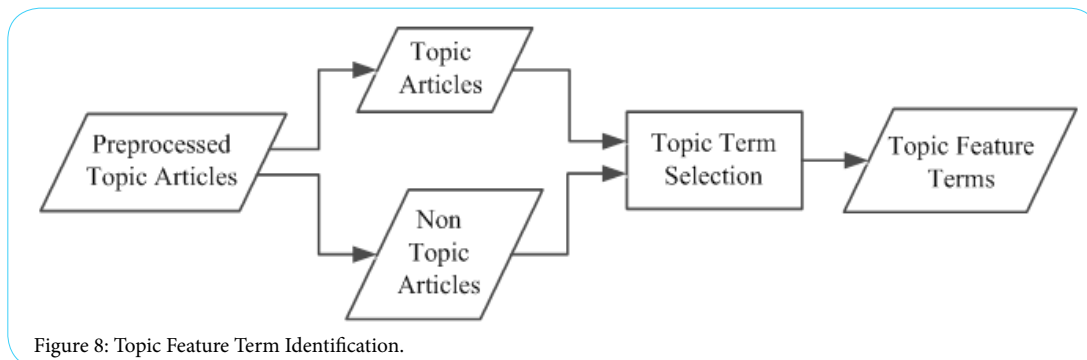

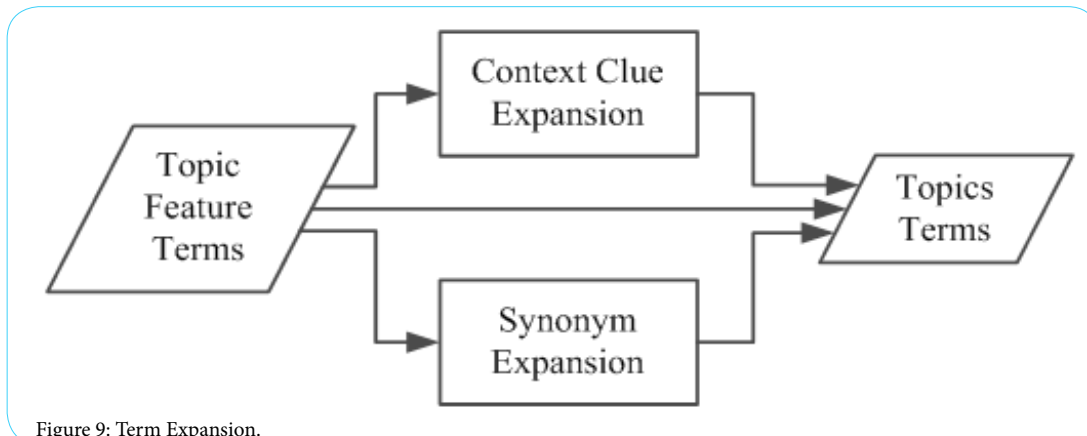Figure 8: Topic Feature Term Identification.

Figure 9: Term Expansion.

Synonym expansion: Synonyms of the terms selected by TF-IDF can be determined using the online Chinese dictionary for artificially tagged synonyms [22], as screened in (Figure 10).

The synonym dictionary is used to identify synonyms related to the topic term. The number of occurrence for these synonyms is counted and those that are the same as stop words are removed. Finally, a threshold is set to keep the synonyms whose occurrence number exceeds the threshold.

eWOM identification

eWOM identification contains topic confirmation for sentences and polarity analysis, as described below.

Topic confirmation for sentences

Topic confirmation for sentences includes the two steps of eWOM preprocessing and topic confirmation, as described as follows.

eWOM Preprocessing

This step preprocesses unknown Internet Chinese articles to obtain meaningful term units. The details in eWOM preprocessing are the same as Internet article preprocessing presented before (Figure 2).

Topics confirmation

Topic confirmation mainly selects sentences conforming to the products of the target enterprise to enhance the accuracy of Chinese eWOM identification. Figure 11 shows the detailed steps involved in topic confirmation.

Sentence segmentation: The article preprocessed in step (1) is segmented based on the sentence unit, and all segmented sentences in the article are outputted sequentially.

Topic term matching: Based on the results in step (a), each sentence is compared with the topic term library. The sentence with the topic term and the following sentence are regarded as the topic sentence.

For example, the topic term "oil content" appears in the sentence "the oil content of snacks is reduced by 70%". This sentence and the subsequent sentence "to eat healthier" thus are considered the topic sentences conforming to the products of the target enterprise.

Polarity analysis

Polarity analysis includes sentence POS-combination classification, sentence polarity analysis, and term polarity analysis. They are explained below.

Sentence POS-combination classification

As shown in Figure 12, the POS-combination rules are determined from the library of POS-combination rule according to the topic sentences analyzed before.

Based on the POS-combination of topic sentences, the most similar POS-combination rules are determined from the library if POS-combination rule via dynamic programming (Equation 3) [16,18]. If POS-combination rules have high similarity and exceed the threshold, they are regarded as the POS-combination rules of the topic sentence. If their similarity is below the threshold, the topic sentence is considered to lack suitable POS-combination rules.
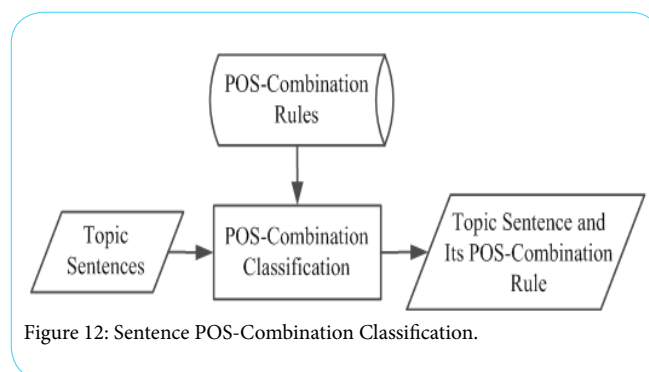

Figure 12: Sentence POS-Combination Classification.

Figure 10: The Online Chinese Dictionary for Artificially Tagged Synonyms.
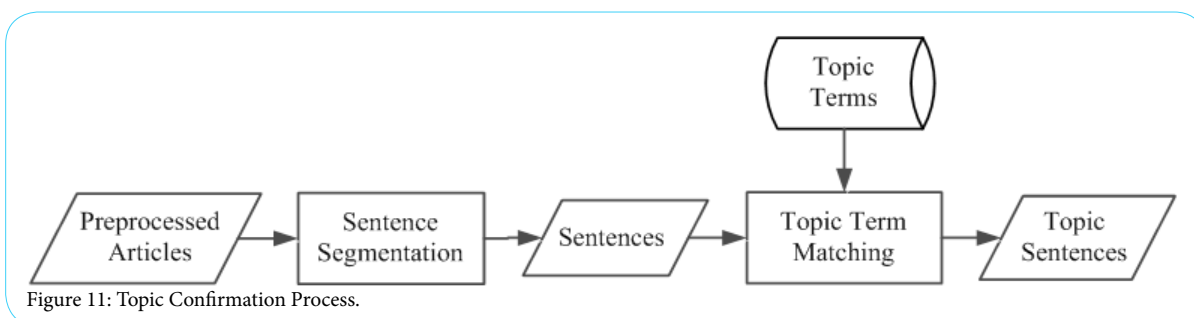

Figure 11: Topic Confirmation Process.

Sentence polarity analysis

The topic sentences and their POS-combination rules acquired in step (1) are processed with the library of polarity term to derive the polarity of the topic sentence, as shown in Figure 13.

Analysis by POS-combination: The location of the polarity term in a sentence is first judged using the POS-combination rules. The polarity term is then matched with the library of polarity terms to obtain its polarity. Finally, the polarity of the term is combined with POS-combination rules to determine the polarity of the topic sentence. For example, the Chinese topic sentence "(這樣)(的)(美食)(能)(錯過)(嗎)" has the POS tags "Vi", "T", "N", "ADV", and "Vt", respectively. POS-combination rules are determined using the libraryof POS-combination rule. The fifth term "錯過" in the Chinese topic sentence is a polarity term, and the sentence meaning is opposite to the polarity term. The term "錯過" is thus searched in the library of polarity term and identified as a negative term. Based on this analysis, the topic sentence tends to have a positive meaning.

Analysis by logic rule: Based on the traditional logic rules, the features of polarity term and the antonyms are used to judge the polarity of sentences. This allows all terms in the topic sentence to first be matched with the library of polarity term to obtain the polarity of one or more polarity terms. The front of the polarity term in the topic sentence is then examined to determine whether it includes an antonym. If so, its polarity is considered opposite.

Term polarity analysis

The polarity of unknown polarity terms is artificially analyzed by the knowledge constructor, and is stored in the library of polarity term for the use of Chinese eWOM analysis.

Mechanism Implementation with an Actual Example

Based on the proposed techniques on POS-combination supported Chinese eWOM analysis, a prototype of POS-combination supported Chinese eWOM analysis is implemented at the Enterprise Information Systems Research Lab (EISRL) of National Kaohsiung First University of Science and Technology, Kaohsiung, Taiwan, ROC. The computer hardware used in the implementation environment is equipped with an application server, web server, and data and knowledge server. The application programs are implemented using Python v2.7.3 and PHP Script language as the development tool, and MySQL 5.0 version as the database.

To examine the feasibility of the proposed techniques, the famous food company "Yu Jan Shin", located in Taichung, Taiwan, is adopted as a real world example for test application of the implemented mechanism. Figure 14 shows an Internet article (eWOM) retrieved from the Chinese Blog and its preprocessing, while Figure 15 lists the results of eWOM identification.

Table 2 lists the accuracy of eWOM identification by POS-combination rules and traditional logic rules. The comparison clearly shows that the overall accuracy of POS-combination rules proposed herein exceeds that of traditional logic rules.

Conclusions

This work developed a suitable technology for POS-combination supported Chinese eWOM analysis to acquire Chinese eWOM sentences for specific target products. Moreover, this study uses POS-combination rules to support the polarity analysis of Chinese eWOM and thus effectively help an enterprise rapidly and accurately understand the current situation of eWOM to improve the customer relationship. The main results and contributions of this work are summarized below.

POS-combination supported Chinese eWOM analysis process: For Chinese eWOM, the POS-combination supported Chinese eWOM analysis process is designed to help an enterprise analyze current Chinese eWOM.

POS-combination rule generation and topic term extraction method: This study used the POS-combination to propose the method of POS-combination rule generation and topic term extraction for the development of eWOM identification.

eWOM identification method: The POS-combination rules and the topic term library are combined to develop a method of eWOM identification that can constantly update the features of eWOM identification and thus effectively enhance the accuracy of eWOM analysis.
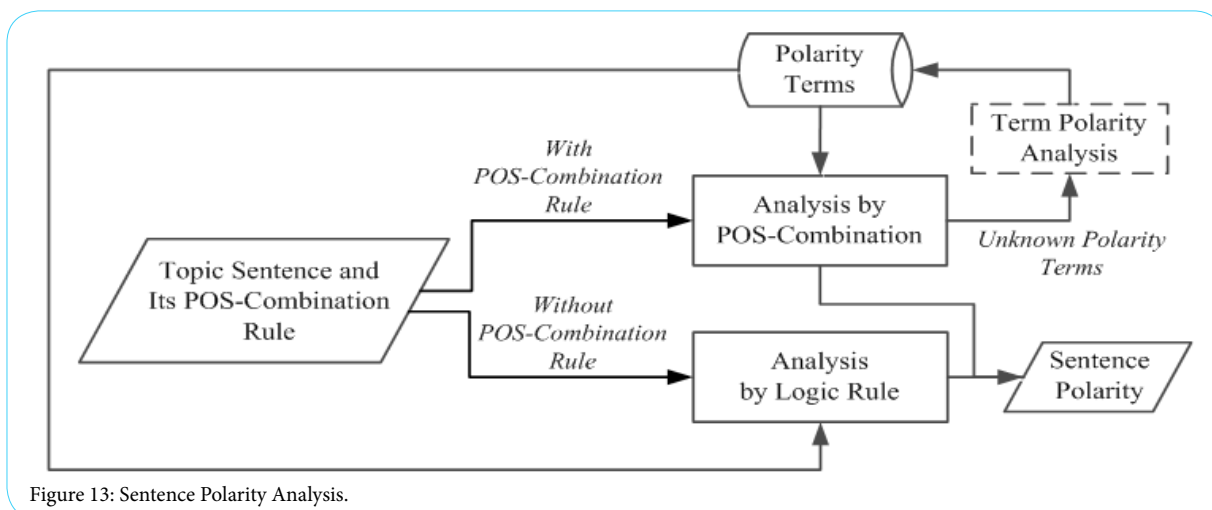

Figure 13: Sentence Polarity Analysis.

Figure 14: Snapshot of an Internet Article and Its Preprocessing.

Figure 15: Snapshot of eWOM Identification by POS-Combination Rules.

Method	Misjudged Sentences	Total Judgment Sentences	Accuracy
POS-Combination Rule	223	3756	94.06%
Traditional Logic Rule	389	4338	91.03%

Table 2: Comparison of eWOM identification Accuracy.

The POS-combination supported the Chinese eWOM analysis mechanism: The mechanism can immediately and accurately analyze current Chinese eWOM to serve as a reference in customer relationship management. Moreover, it can provide the basis of future research and applications in eWOM.

Acknowledgements

Competing Interests

The authors declare no competing interest.

References

1. Broden TF (2006) Paradigm versus Syntagm, Encyclopedia of Language & Linguistics.

2. Cantallops AS, Salvi F (2014) New consumer behavior: a review of research on eWOM and hotels. International Journal of Hospitality Management 36: 41-51.

3. Chen K, Ma W (2002) Unknown word extraction for Chinese documents. Proceedings of COLING 1: 1-7.

4. Chen Y, Shang R, Li M (2014) The effects of perceived relevance of travel blogs' content on the behavioral intention to visit a tourist destination. Computers in Human Behavior 30: 787-799.

5. Darong H, Peng W (2012) Grid-based DBSCAN algorithm with referential parameters. Physics Procedia 24: 1166-1170.

6. Edla DR, Jana PK (2012) A prototype-based modified DBSCAN for gene clustering. Procedia Technology 6: 485-492.

7. Fang W, Roberts FS, Ma Z (2001) A measure of discrepancy of multiple sequences. Information Sciences 137: 75-102.

8. Feng S, Pang J, Wang D, Yu G, Yang F, et al. (2011) A novel approach for clustering sentiments in Chinese blogs based on graph similarity. Computers & Mathematics with Applications 62: 2770-2778.

9. Friederici AD, Weissenborn J (2007) Mapping sentence form onto meaning: the syntax-semantic interface. Brain Res 1146: 50-58.

10. Ghiassi M, Skinner J, Zimbra D (2013) Twitter brand sentiment analysis: a hybrid system using n-gram analysis and dynamic artificial neural network. Expert Systems with Applications 40: 6266-6282.

11. He Y, Zhoub D (2011) Self-training from labeled features for sentiment analysis. Information Processing & Management 47: 606-616.

12. Lo S (2008) Web service quality control based on text mining using support vector machine. Expert Systems with Applications 34: 603-610.

13. Ma W, Chen K (2003) A bottom-up merging algorithm for Chinese unknown word extraction. Proceedings of ACL workshop on Chinese Language Processing 17: 31-38.

14. Pai M, Chu H, Wang S, Chen Y (2012) Electronic word of mouth analysis for service experience. Expert Systems with Applications 40: 1993-2006.

15. Quan C, Ren F (2010) A blog emotion corpus for emotional expression analysis in Chinese. Computer Speech & Language 24: 726-749.

16. Rong A, Figueira JR (2013) A reduction dynamic programming algorithm for the bi-objective integer knapsack problem. European Journal of Operational Research 231: 299-313.

17. Singh T, Veron-Jackson L, Cullinane J (2008) Blogging: A new play in your marketing game plan. Business Horizons 51: 281-292.

18. Sitarz S (2014) Multiple criteria dynamic programming and multiple knapsack problem. Applied Mathematics and Computation 228: 598-605.

19. Zhai Z, Xu H, Kang B, Jia P (2011) Exploiting effective features for Chinese sentiment classification. Expert Systems with Applications 38: 9139-9146.

20. Zhang M, Fang W, Zhang J, Chi Z (2005) MSAID: multiple sequence alignment based on a measure of information discrepancy. Comput Biol Chem 29: 175-181.

21. https://books.google.com/ngrams

22. http://www.kwuntung.net/synonym/