American Journal of Applied Sciences

Original Research Paper

Opinion Extraction on Online Malay Text

Surendran Selvaraju and Kalaiarasi Sonai Muthu Anbananthen

Faculty of Information Science Technology, Multimedia University, Malacca, 75450 Malaysia

Article history
Received: 22-10-2018
Revised: 11-01-2019
Accepted: 10-06-2019

Corresponding Author:
Surendran Selvaraju
Faculty of Information Science Technology, Multimedia University, Malaysia
Tel: +606-2523344
Fax: +606-2318840
Email:surendranmmu@yahoo.com
kalaiarasi@mmu.edu.my

Abstract: Growing of social media usage present a new set of opportunities and challenges in the way of information is retrieved and searched. Opinions on social media has become an important factor in influencing people choices on purchasing a product and service. Hence, sentiment analysis has become the most crucial tool in tracking people feedbacks on products and services. For Malay language there is limited sources available for this language. Thus, in this paper we present the method of extracting opinion on online Malay text. The traditional method using POS extraction is not adequate. Thus, rule based method is integrated with POS extraction method to improve opinion words extraction. Most of the existing tools are able to retrieve opinion at sentence and document level. More detail analysis is acquired to have detail information and summarization of a product. This is where feature level sentiment analysis is needed. The process of identifying opinion of a particular feature in a sentence, can be quite tedious and troublesome. This is because opinion of the feature can be hidden and scattered in the sentence. Therefore, opinion mapping is employed for opinion extraction at feature level in this paper. A set of tweets from telecommunication domain is used to evaluate the proposed framework. From the experiment, the accuracy of the extraction performed is 88%. The detail description of the feature level opinion extraction steps is discussed in this paper.

Keywords: Sentiment Analysis, Opinion Word, Malay Online Text, Feature Level Extraction

Introduction

Social media has seen a steady increase of its usage over the past few years. People use social media as a platform to share their feedbacks and opinions, many of which are easily viewed by the public. This has inadvertently generated a gigantic amount of data online. Amongst many others, this gigantic data also contains customer reviews and feedbacks on various products and services. These reviews and feedbacks plays an important role in decision making. Online customers depend on these reviews and feedbacks before deciding to purchase a product. These valuable opinions are able to easily influence the decision of potential customers. With all these opinion data generated online, businesses too have realized the importance of gathering a customer's feedback database which would prove useful for businesses to plan their marketing and product development.

About 85% of Malaysians are using social media (San *et al*., 2015). About 55% of these users use Malay language to comment and give feedbacks online (San *et al*., 2015). Malay language is spoken not only in Malaysia but it's also used in countries like Brunei, Singapore, Indonesia, Philippines, Central Eastern Sumatra Riau islands and Thailand. This adds up to about 270 million users of this language. There are a lot of reviews, comments and feedbacks in Malay language, however very limited research is done for Malay language in sentiment mining (Samsudin *et al*., 2013). This has led to the difficulty in analysing online Malay text. It is tedious and time-consuming for any individual or businesses to scheme through and capture opinions of these online Malay texts. Therefore, it is crucial to develop a sentiment mining tool which able to analyse the sentiment in Malaysian context.

Sentiment analysis is a growing research area which mainly focus on knowledge discovery and information retrieval from text using natural language processing techniques (Liu, 2012). The goal of sentiment analysis is to enable computer to understand and track emotions expressed online. Sentiment can be defined as a view, thought or feeling. Therefore, sentiment analysis is sometimes called as opinion analysis. Business organizations are investing a large sum of money through surveys and consultation to find out what customers feels

about their products. Additionally, individuals such as business owner or marketer are also interested in tracking opinion about their issues, services, products and events so that they able to improve their services and attract more customers.

The automatic extraction and analysis of those messages posted in online site becomes a popular research topic in the recent years. Sentiment analysis or information extraction at the feature level is needed as more detail information need to be extracted. In the task of sentiment mining has been only focusing the overall polarity of a text. Indeed, with this feature level information people would able to acquire a better and clear information on products or services. The task of extracting opinion at feature level can be technically challenging but very useful in practice. This paper aim to extract feature level opinion of products and services on online Malay text. A novel method has been introduced to carry out the extraction of opinion on online Malay text.

Work Related

In this session, different methods of opinion extraction are reviewed and summarized as below.

Opinion extraction supports various tasks such as sentiment analysis of reviews, document classification and insight summarization (Wicaksono and Myaeng, 2013). The task of extracting opinions at feature level is very challenging, however it has proven to be beneficial in providing detailed insights on products. To further expound, some people might like the services provided by a hotel, but some people might also only focus on certain features of the hotel such as food and decorations. Furthermore, it can be tedious work extracting opinions from online texts. This is because sentences can sometimes be written while completely disregarding rules of grammar.And since the sentences are grammatically incorrect, the opinions of the feature are scattered all over the sentence. There are three main methods as shown in Fig. 1, that can be used to extract opinions - (1) POS tag (2) Rule-based and (3) Distance.

POS Extraction

POS extraction method using part of speech to extract word in a sentence. POS extraction is important and necessary to determine the features and opinion words in a sentence. Sharma *et al*. (2014) used POS tag to identify features and opinions in customer reviews of mobile phones, while Htay and Lynn (2013) used this extraction method to extract opinion words that described the features of a product from review texts. The POS tags used to identify opinion words are adverbs and adjectives. Adverbs and adjectives that are close to the feature are regarded to as opinion words, the features in the review are then extracted using nouns. Thereafter, the features extracted are used to locate the opinion words in the review, along with the idea that the opinion word is closest to the feature. Zhang and Liu (2011) have also utilized POS tag to find opinion words in review dataset. The focus is on nouns that imply opinions. In Malay POS tag, nouns (kata nama) is regarded as feature (Alsaffar and Omar, 2014). To conclude, POS extraction is able to extract all the opinion words in a sentence but often results in extracting unwanted words too.

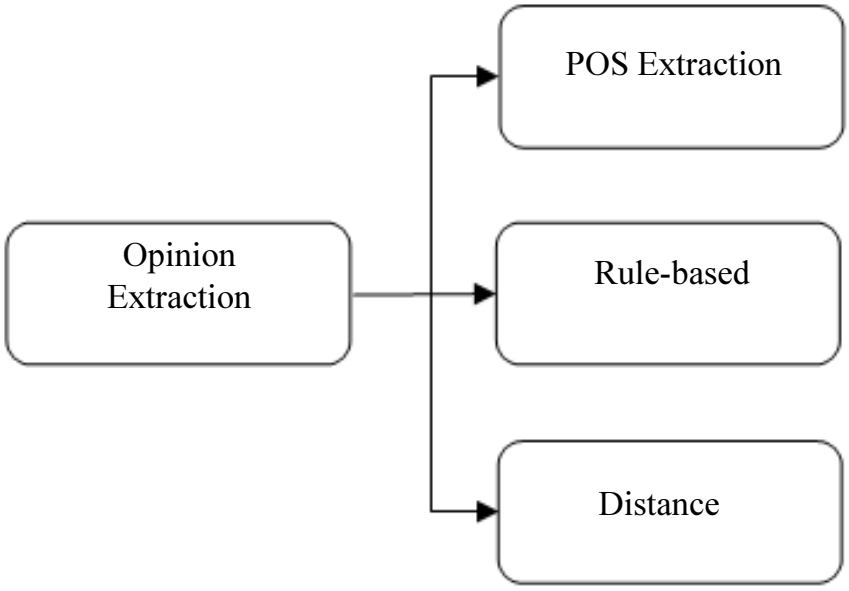

Fig. 1: Opinion extraction methods

Rule-Based

Rule-based works based on syntactic rule in a sentence. Peng and Shih (2010) have used the rule based method to extract opinion words on document and sentence levels. The polarity of the words is obtained by referring to the sentiment lexicon. And for opinion words which are not found in the lexicon, their polarity is determined by referring to words with closest meanings and known polarity. Qiu *et al*. (2011) used the double propagation method through dependency to extract potential opinions and potential opinion targets (feature). For instance, in an opinion sentence "Nokia takes awesome pictures!", the word awesome is parsed as directly reliant on the noun picture. If the word "awesome" is recognized as an opinion word, the picture can then be extracted as target. Similarly, if the picture is identified as target, then the word "awesome" can be extracted using the same rule. A similar method is used by (Cruz *et al*., 2010; Kumar and Raghuveer, 2012; Golpar-Rabooki *et al*., 2015) to extract opinions and features in reviews. Babu and Das (2015) used syntactic dependency to identify opinion words in a sentence. The Stanford dependency parser is used to get the dependency relations between opinion words and features. The feature in the sentence is identified using feature dictionary.

Later on, the opinion words are used to calculate the polarity of features in product reviews. Zhou *et al*. (2014) studied feature-opinion extraction from online reviews. The features and opinions in the reviews are extracted using rule-based. A rule sets is developed to capture the general patterns that are used by customers when expressing their opinions. In general, rule based method works well in grammatically correct sentences. The disadvantage of this method is that it works poorly on online text where rules of linguistics are not followed.

Distance

Normally, the product features and opinion word is not independent from each other. The opinion word that are used to describe the feature would always be located around the feature in a sentence. Based on observations made, it can be learned that opinion words can be extracted by extracting the adjective and verb near the features (Godbole *et al*., 2007; Somprasertsri and Lalitrojwong, 2010). Popescu and Etzioni (2007), OPINE which is an unsupervised information system is used to extract important product features in reviews. The system is able to identify features and opinions regarding the features, as well as to determine the polarity of opinions. The system recursively identifies features until no other candidates are found. KnowItAll, a web based domain-independent information extraction system is used by OPINE to generate extraction rules to extract opinions. Frequent features are used to identify opinion candidates. This method is carried out with the assumption that opinion words (adjective only) associated with product features are always nearby. Ding *et al*. (2008) used the same ideology to determine the polarity of opinions expressed in reviews on a product's feature. The feature in a review sentence is detected using POS tag. The opinion words in the sentence are then identified and the polarity of the words are retrieved using a lexicon. The distance method is used to compute the sentiment score for feature. Marrese-Taylor *et al*. (2014) have used the same method to determine the features polarity of tourism products. The distance method works well on documents and reviews as the opinion word is always written near to feature. However, this method does not work well for online texts. The opinion word extracted from online texts may be incorrect and this is because opinion words on online texts aren't always near to feature, for it can also be written far away from the feature. The distance method is effective and fairly simple for detection of opinions on a feature, however this method presents a disadvantage whereby extracting opinion words that is nearby to the feature can be incorrect. A sentence may also express opinions on multiple features (Popescu and Etzioni, 2007).

Methodology

In this session the opinion extraction methodology is explained in detail.

Opinion extraction is basically a very crucial step in determining where in a sentence the opinion of a feature is embedded. Opinions can be extracted from a feature, sentence and at document level. For this research, extraction of opinion at feature level will be more focused upon. Opinion words convey either positive or negative polarity. Initially, the words tagged with "KA" (adjective) and "KK" (verb) are believed to be opinion words:

"Telekom (KNK) punya (KT) wifi (KN) perlahan (KA) kakak (KN) marah(KA**)"**

From the sentence, both the words "perlahan" and "marah" will be extracted as opinion words. Despite, that "marah" is not the opinion expressed for the feature "wifi".

Based on observations made, it is discovered that not all the words that have been tagged as adjective or verb are useful enough to be considered as opinion words. A certain number of the extracted opinion words are unwanted words that somehow fall under the category of opinion words. And this is what makes most situations difficult, as most online sentences have usage of inappropriate and unnecessary words in them and they usually don't contribute to any meaning as opinion words. A lot of these extracted opinion words are not used to express opinions for the desired feature in a sentence. Hence, an opinion extraction framework is proposed to improve the extraction process.

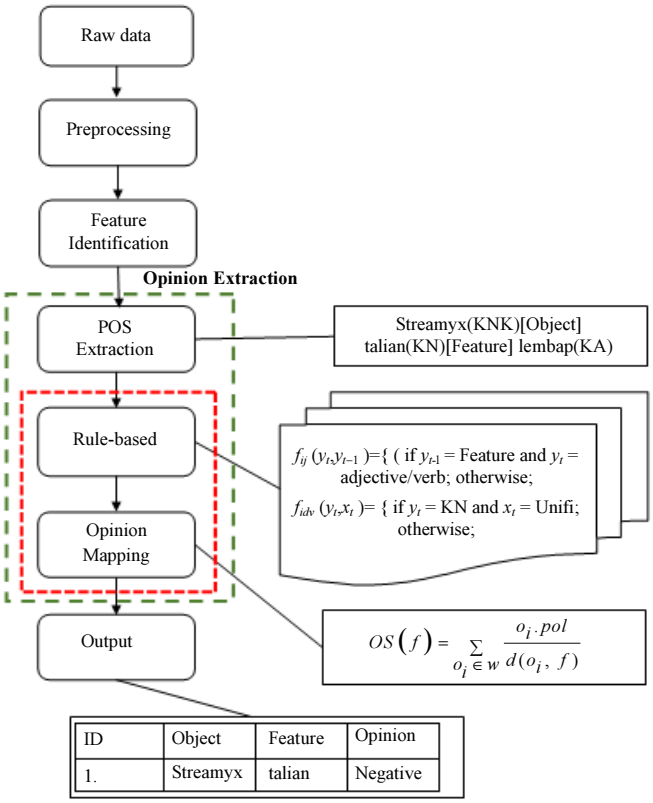

Fig. 2: Opinion Extraction Framework (PRO)

Figure 2 shows proposed framework for the opinion extraction. This opinion extraction framework is called PRO. And in this framework, rule based and opinion mapping are introduced along with POS extraction method. A detailed explanation of each method will be described below. The POS extraction method is the same as described initially, whereby the opinion words in the sentence are extracted based on "KA" and "KK" POS tags. So basically, rule-based and opinion mapping methods are integrated to improve the traditional POS extraction method.

Rule-Based

The Rule-Based Method is employed to improve the opinion word identification and extraction. The rules are devised by observing the words and the labels at position t, $t+1$ and $t-1$. In rule-based, the general form of rule is $f_{idv}(y_t, x_t) = y_t$ = POS and x_t = word, which looks at a pair of adjacent states x_{t-1} and x_t. y is the POS tag label and x is the observation word. Figure 3 shows some sample rules derived from online Malay sentences. Rule 1: If the previous POS tag is "KT" (else) and the current POS tag is "KA" (adjective), then the current word should be considered an opinion word. Rule 2: If the previous label is Feature and the current word POS tag is "KA" (adjective), then the current word should be extracted as an opinion word. This Rule-Based method has significantly reduced the extraction of unwanted opinion words. Although the opinion word candidates have been reduced, however there is still a problem in handling mixed opinions in a sentence. This is simply because the opinion words don't express opinions for the desired feature that was extracted. For example:

"Streamyx (KN) bagus (KA) dekat (KT) syarikat (KN) bodoh (KA) ini (KT)".

From the given sentence example, "bagus (KA)" and "bodoh (KA)" will be extracted as opinions for the feature "Streamyx". Even though the word "bodoh" doesn't express opinion for the feature "Streamyx", it is still expressing opinion for the feature "syarikat". It certainly is a challenging task to analyse opinion at feature level, especially with informal text language used in blogs and tweets, as these sentences are full of grammatical errors. Besides this, mapping the correct opinion word to a particular feature is also very challenging. Therefore, solely using the POS-tag approach in extracting opinion words has proven to be inadequate. For example:

$f_i(y_{t-1}, y_t)$ = {if y_{t-1} = kata tugas and y_t = KA/KK; otherwise;}	if the previous POS tag else (kata tugas) and the current word POS tag is KA/KK, then the current word should be considered as opinion word.
$fi(y_{t-1}, y_t)$ = {(if y_{t-1} = Feature and y_t = KA/KK; otherwise;}	if the previous word label is feature and the current word POS tag is KA/KK , then the current word should be considered as opinion word.
$f_i(y_{t-1}, y_t)$={(if y_{t-1} = KN/KNK , y_t = KA/KK and $y_t \neq$ opinion; otherwise;}	if the previous word POS tag is KN/KNK and the current word POS tag is KA/KK, then the current word should be considered as opinion word.
$f_i(y_t, x_{t-1})$= {if x_{t-1} = sangat/selalu and y_t = KA/KK; otherwise;}	if the previous word is "sangat/selalu" and the current word POS tag is KA/KK, then the current word should be considered as opinion word.
$f_i(y_{t+1}, y_t)$= { if y_{t+1} = N, y_t = KA/KK and $y_t \neq$ opinion ; otherwise;}	if the word after has the tag N and the current word POS tag is KA/KK, then the current word should not be considered as opinion word

Fig. 3: Rules of opinion extraction

"Unifi(KNK) punya(KT) wifi(KN) dalam(KA) laju (KA)"

"Unifi (KNK)" is the object and "wifi(KN)" is the feature for the sentence. The opinion word expressed for the feature is "laju (KA)", but since the word "dalam" also has POS tag of "KA", it will be also extracted as one of the opinion word which is not true for the sentence given. Here's another example:

"Telekom(KNK) punya (KT) talian (KN) dekat (KA) sini (KT) bagus (KA)"

In the sentence above, "Telekom(KNK)" is the object and "talian(KN)" is the feature for the object. The word "bagus(KA)" is the opinion expressed for the feature. POS tag of the word "dekat" is also "KA". Hence, "dekat(KA)" will be also extracted as one of those opinion words in the sentence. Even though rule-based has reduced the opinion, however mapping still becomes a problem. Therefore, in the next section, opinion mapping will be introduced. In this research, the opinion mapping method is integrated to map the correct opinion to a feature.

Opinion Mapping

The method used in opinion mapping is called Opinion Score (OS). OS is used to determine the opinion polarity of a feature in a sentence, which is adapted from Liu and Zhang (2012). In a given sentence, the OS of a feature is calculated. Positive opinion words (Anbananthen *et al*., 2017) are assigned the polarity score of +1 and negative opinion words polarity score of −1. All the polarity score of the feature will be summed up by using the following score function as shown by Equation 1.0:

$$OS(f) = \sum_{o_i \in w} \frac{o_i.pol}{d(o_i, f)} \quad (1.0)$$

O_i is an opinion word, whereas w is the set of all the opinion words in the sentence. The $d(O_i, f)$ in the equation translates as the distance between the opinion word and the feature in the sentence. The multiplicative inverse in the Equation 1.0 is to provide low-weight age to opinion words that are a distance from the feature *f*.

The final score obtained from the sum of all the opinion words in the sentence would determine the polarity of the feature. The polarity of the feature in the sentence will be deemed positive if the final polarity score is positive and likewise, it would be negative if the final polarity score is deemed negative. The polarity scoring works with the assumption that - opinion words that is far from the object/feature may not be the opinion for the object/feature:

$$d(X,Y) = \sqrt{\sum_{n=t}^{n} (\Delta x + \Delta y)^2} \quad (1.2)$$

The $d(O_i, f)$ is calculated using Euclidean distance formula which is shown in Equation 1.2, where *i* is the position of opinion word in the sentence. For instance:

"Teruk(KA) speed (KN) Unifi(KNK) ini(KT) tapi(KT) emak(KN) gembira(KA)"

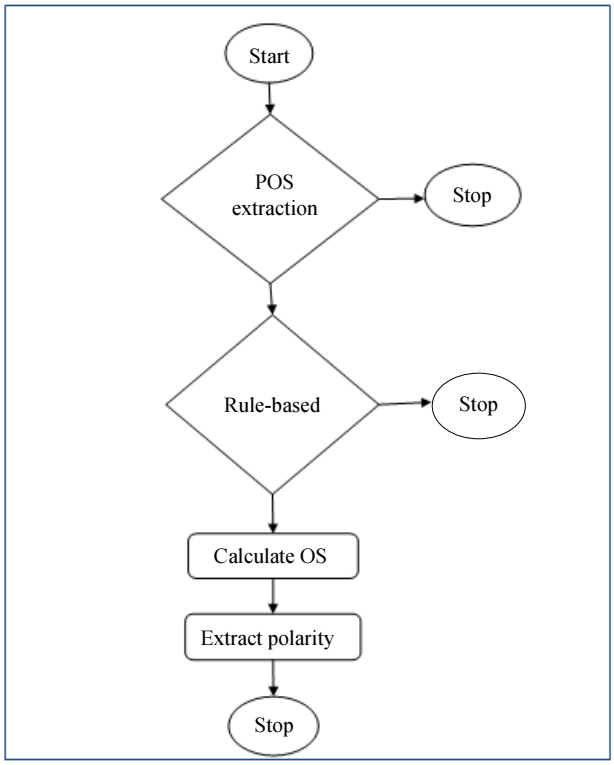

Fig. 4: Opinion extraction flowchart

Based on the sentence, Unifi is identified as the object while speed is identified as the feature. There are adjective words in the sentence and they both have different polarity - "teruk" is negative while "gembira" is positive. Since both these adjective words satisfy the rules for opinion words selection, the *OS* score of each of the adjective words from the feature will be computed as follows.

"Teruk" is negative (-1) and the distance from the feature is 1, so it has $OS = -1$, while "gembira" is positive (+1) and the distance from the feature is 5, so it has $OS = 0.2$. The two *OS* is summed up to give the final value *OS* of the feature which is -5. Since the final *OS* <1, the polarity of the feature will be deemed negative.

From the flowchart in Fig. 4, the input of the system will be tagged as shown in the sentence above. Based on the tagged sentence, the opinion words are then identified. Each position of the word *t* will be checked for verb and adjective tagging. Only the verbs and adjectives that satisfy the rules will be considered as opinion words. Once the word is considered opinion words, the score of the opinion words against the opinion target (feature) will be computed.

The Opinion Score (OS) of each opinion in the sentence will be calculated and summed up to determine the final score. If the final score is greater than >1, the feature would then have positive polarity. If the final score is less than <1, the feature would have negative polarity. helped in mapping the correct opinion polarity for the feature.

Experimant and Discussion

This section evaluates the automated extraction of opinion from online Malay text at feature level. 500 tweets were used to test the extraction process. Total of 1515 opinions words are extracted from the 500 tweets using POS extraction method. The 1515 opinion words are further refined using the Rule-Based method. Through integration of this method in opinion extraction, the number of opinion words have been reduced from 1515 to 890. Through these combined methods, a total of 625 words were reduced from the extraction. It can also be observed that certain words extracted as opinion words using the POS method aren't actual opinion words for the feature. But when the rule-based module is integrated, the extraction process of opinion words was further improved.

Table 1 shows the comparison of opinion words extracted using the POS extraction method and the combination method; which is POS Extraction +Rule-based method. As can be seen, Column 1 shows the extracted opinion words with POS extraction method, whereas Column 2 displays the opinion words extracted using the combination method (POS extraction + Rule-based).

Table 1: Opinion words extracted

POS extraction	POS extraction + Rule-based
baru (Adjective)	bahagia (Adjective)
bahagia (Adjective)	lembap (Adjective)
lembap (Adjective)	masalah (Adjective)
habis (Adjective)	ada (Adjective)
masalah (Adjective)	ada (Adjective)
baik (Adjective)	pasang (Verb)
ada (Adjective	percuma (Adjective)
lepas (Adjective)	marah (Adjective)
ada (Adjective)	jaga (Verb)
pasang (Verb)	betulkan (Verb)
percuma(Adjective)	baik (Adjective)
marah(Adjective)	pakai (Verb)
jaga (Verb)	susahkan (Adjective)
betulkan (Verb)	ok (Adjective)
baik (Adjective)	ada (Adjective)
pakai (Verb)	laju (Adjective)
susahkan (Adjective)	perlahan (Verb)
ok (Adjective)	baru (Adjective)
ada (Adjective)	
cuba (Verb)	
masuk (Verb)	
laju (Adjective)	
perlahan (Verb)	
minta (Verb)	
beli (Verb)	
baru (Adjective)	

In this table, it is clearly indicated that the number of opinion-word candidates have been reduced in number. Some words extracted as opinion words using the POS extraction method in Column 1 are not extracted when the Rule-based is applied in the extraction process, as shown in column 2.

Table 2 shows the result of opinion extraction using only POS extraction. The extracted opinion-word candidates are validated manually. From the table it can be seen that the false negative is 0% while false positive is about 42%, which is 625 words. As discussed earlier this method is resulted in extracting all the unwanted words as opinion words. Not all the words tagged with “KA” and “KK” are opinion words expressed for features in a sentence. This is further explaining by the precision reading, which is 43%. Eventhough, the recall read is 100% which is able to extract all the opinion words but the accuracy is only 45% since it is extracting unwanted words as well.

Results of performance of the opinion extraction process using the combination of POS extraction and Rule-based method is shown in Table 2. The extracted opinion-word candidates are validated manually. From the validation, a confusion matrix table is produced to study the results as tabulated above. As seen, the false-positive reading is 8% of the total extracted opinion-word candidates, which means that 70 words aren’t opinion words, but they were extracted in the process.

Table 2: Result of opinion extraction using POS extraction

	Opinion word	Not opinion word	Total
Predicted opinion word	470	625	1095
Predicted not opinion word	0	420	420
Total	470	1045	

Accuracy = (TP+TN)/Total = (470+420)/1515 = 0.45
Recall = TP/ TP + FN = 470/470+0 = 1.00
Precision = TP/ TP + FP = 470/470+625 = 0.43

Table 3: Result of opinion extraction

	Opinion word	Not opinion word	Total
Predicted opinion word	370	70	440
Predicted not opinion word	100	350	450
Total	470	420	

Accuracy = (TP+TN)/total = (370+350)/890 = 0.81
Recall = TP/ TP + FN = 370/ (370+100) = 0.78
Precision = TP/ TP + FP = 370/ (370+70) = 0.84

Table 4: OS results

	Actual Positive	Actual Negative	
Predicted positive	260	20	290
Predicted negative	40	180	210
	300	200	

Accuracy = (TP+TN)/total = (260+180)/500 = 0.88
Recall = TP/ TP + FN = 260/ (260+40) = 0.87
Precision = TP/ TP + FP = 260/ (260+20) = 0.93

This is because the extracted words match the designed rules. . However, they don’t act as opinion for the features, but instead act as opinions for other entities.

For example, in the sentence: "awak (KN) ini (KT) bangga (KA) tapi (KT) talian (KN) bodoh (KA)", the word “bangga” will be extracted as one of the opinion words as it matches the first rule. The first rule states that if the previous POS-tag (kata tugas) and the current POS-tag is KA/KK, then the current word should be considered an opinion word. Even though the word “bangga” matches the rules and is extracted as an opinion word, this does not imply an opinion for the feature (talian), but instead an opinion for an entity (awak). From the result, the false-negative stands at 11%, which equals 100 opinion words predicted to be not opinion words. Some of the opinion words satisfy the rules designed, which is to not be extracted as opinion words. An example of the rule is as such - If the previous word POS tag is “N” and the current POS tag is KA/KK, then the current word should not be considered as an opinion word. The reason something like this would happen is due to the poor writing habits on Twitter by its users. In this experiment, sentences on Twitter have been used to assist with the experiment, so sentences with

improper grammatical structure is expected. From the sentence- “wifi (NOUN) kura (NOUN) perlahan (ADJ)", even though, the word "perlahan" brings opinion for "wifi" but from the rules designed, it is assumed the opinion is for the word "kura". From the confusion matrix shown in Table 3, it can be concluded that Rule-based method does improves the identification and extraction of opinion words. The accuracy of the extraction is 81%, while the recall and precision of this approach are 78% and 84% respectively.

In order to solve the feature opinion extraction, opinion mapping is introduced as discussed in the methodology. The experiment is then carried out by adding opinion mapping module to POS extraction and Rule-based method to determine the correct opinion polarity for a feature in a sentence. Table 4 shows the result of opinion score. The accuracy of the opinion mapping method is about 88%. The precision and recall are 93% and 87% respectively. From the result, it can be deduced that the opinion score method thus helps in opinion extraction of a feature which is considerably accurate. The false-negative and false-positive percentage are 6% and 5% respectively. This is due to some of the sentences have wrong usage of words which contribute to having one or more “KA/KK” to be present side by side in a sentence. In this sentence for example - “menipu (ADJ) betul (ADJ) servis (NOUN) Streamyx (NOUN)”, the word “menipu” has negative polarity, while the word “betul” is positive. Both of the words opinion score will be calculated and summed up. According to the OS calculation, the word “servis” would be given positive polarity even though it’s supposed to be negative. This is because, both of the opinion words are placed side by side, which technically is wrongly placed and grammatically incorrect. And since the word “menipu” is placed far from the feature, it will have a lower score compared to the word “betul”. Thus, the total score produced would be >1 and the feature will have a positive polarity. Besides this, the wrong spelling of words can also contribute to false-positive and false-negative readings. Wrongly spelt words are tagged as “KT” and the word with POS tag “KA/KK” that comes right after is identified as an opinion candidate as it matches the rule. This somehow leads to wrong opinion word identification. In another example, wrongly spelt opinion words are tagged as “KT” instead of “KA/KK” and this causes it to be opted out as opinion word candidates. This then results in false-positive or false-negative polarity extraction when the OS calculation is performed.

Conclusion

The usage of social media in Malaysia is hiking up. Most of the Malaysian are using social media as a platform for sharing and providing feedbacks on products and services. Many sentiment tools available are not capable of processing Malay online text and in addition to it, majority of the tools are developed for document and sentence level but none for feature level. Indeed, feature level sentiment analysis provide more detail information of a product. Therefore, in this paper Malay opinion extraction is proposed and developed for Malay language at feature level. Besides, there are fewer resources for Malay language in sentiment analysis to be referred. Hence, the framework for opinion extraction has been carefully studied and developed. The framework is developed based on Malay online text which is differ from Malay document text. Malay online text is considered as improper grammar text. Based on our experiment results, the performance of the proposed opinion word extraction and opinion mapping method is considerably acceptable. In future, this proposed technique can be enhance incorporating machine learning techniques for automated opinion word rules generation which will boost the extraction accuracy.

Acknowledgements

This work has been supported by TM R&D Grant No. MMUE/140065.

Author’s Contributions

Surendran Selvaraju: Introduction, Literature review, Methodology, Experiment and evaluation, Conclusion.

Kalaiarasi Sonai Muthu Anbananthen: Introduction, Methodology, Experiment and evaluation.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

References

Alsaffar, A. and N. Omar, 2014. Study on feature selection and machine learning algorithms for Malay sentiment classification. Proceedings of the Proceedings of the 6th International Conference on Information Technology and Multimedia, Nov. 18-20, IEEE Xplore Press, Putrajaya, Malaysia, pp: 270-275. DOI: 10.1109/ICIMU.2014.7066643

Anbananthen, K.S.M., S. Selvaraju and J.K. Krishnan, 2017. The generation of malay lexicon. Am. J. Applied Sci., 14: 503-510. DOI: 10.3844/ajassp.2017.503.510

Babu, S.M. and S.N. Das, 2015. An unsupervised approach for feature based sentiment analysis of product reviews. Int. J. Sci. Res. Eng. Technolo.

Cruz, F.L., J.A. Troyano, F. Enríquez, F.J. Ortega and C.G. Vallejo, 2010. A knowledge-rich approach to feature based opinion extraction from product reviews. Proceedings of the 2nd International Workshop on Search and Mining User-Generated Contents, Oct. 30-30, IEEE Xplore Press, Toronto, ON, Canada, pp: 13-20. DOI: 10.1145/1871985.1871990

Ding, X., B. Liu and P.S. Yu, 2008. A holistic lexicon-based approach to opinion mining. Proceedings of the 2008 International Conference on Web Search and Data Mining, Feb., 11-12, IEEE Xplore Press, Palo Alto, California, USA, pp: 231-240. DOI: 10.1145/1341531.1341561

Godbole, N., M. Srinivasaiah and S. Skiena, 2007. Large-scale sentiment analysis for news and blogs. ICWSM, 7: 219-222.

Golpar-Rabooki, E., S. Zarghamifar and J. Rezaeenour, 2015. Feature extraction in opinion mining through persian reviews. J. AI Data Mining, 3: pp: 169-179. DOI: 10.5829/idosi.JAIDM.2015.03.02.06

Htay, S.S. and K.T. Lynn, 2013. Extracting product features and opinion words using pattern knowledge in customer reviews. Sci. World J., 2013: 1-5. DOI: 10.1155/2013/394758

Kumar, R.V. and K. Raghuveer, 2012. Web user opinion analysis for product features extraction and opinion summarization. Int J. Web Semantic Technol., 3: 69-82.

Liu, B., 2012. Sentiment analysis and opinion mining. Synthesis Lectures Human Language Technologies, 5: 1-167.

Liu, B. and L. Zhang, 2012. A survey of opinion mining and sentiment analysis. In Mining text data. Springer US, pp: 415-463.

Marrese-Taylor, E., J.D. Velásquez and F. Bravo-Marquez, 2014. A novel deterministic approach for aspect-based opinion mining in tourism products reviews. Expert Syst. Applicat., 41: 7764-7775. DOI: 10.1016/j.eswa.2014.05.045

Peng, T.C. and C.C. Shih, 2010. An unsupervised snippet-based sentiment classification method for chinese unknown phrases without using reference word pairs. Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, Aug. 31, IEEE Xplore Press, Toronto, ON, Canada, pp: 243-248. DOI: 10.1109/WI-IAT.2010.229

Popescu, A.M. and O. Etzioni, 2007. Extracting Product Features and Opinions from Reviews. In: Natural Language Processing and Text Mining, Kao, A. And S.R. Poteet (Eds.), Springer London, pp: 9-28.

Qiu, G., B. Liu, J. Bu and C. Chen, 2011. Opinion word expansion and target extraction through double propagation. Computational Linguistics, 37: 9-27. DOI: 10.1162/coli_a_00034

Samsudin, N., M. Puteh, A.R. Hamdan and M.Z.A. Nazri, 2013. Immune based feature selection for opinion mining. Proceedings of the World Congress on Engineering (WCE' 13), IEEE Xplore Press, London, U.K. pp: 3-5.

San, L.Y., A. Omar and R. Thurasamy, 2015. Online purchase: A study of generation y in malaysia. Int. J. Bus. Management, 10: 1-7.

Sharma, R., S. Nigam and R. Jain, 2014. Mining of product reviews at aspect level. Int. J. Foundat. Comput. Sci. Technolo.

Somprasertsri, G. and P. Lalitrojwong, 2010. Mining feature-opinion in online customer reviews for opinion summarization. J. UCS, 16: 938-955.

Wicaksono, A.F. and S.H. Myaeng, 2013. Toward advice mining: Conditional random fields for extracting advice-revealing text units. Proceedings of the 22nd ACM International Conference on Information and Knowledge Management, Oct. 27, IEEE Xplore Press, New York, USA, pp: 2039-2048. DOI: 10.1145/2505515.2505520

Zhang, L. and B. Liu, 2011. Identifying noun product features that imply opinions. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Jun. 19-24, IEEE Xplore Press, Stroudsburg, PA, USA, pp: 575-580.

Zhou, E., X. Luo and Z. Qin, 2014. Incorporating language patterns and domain knowledge into feature-opinion extraction. Proceedings of the International Conference on Text, Speech and Dialogue, Springer, Cham, pp: 209-216. DOI: 10.1007/978-3-319-10816-2_26