

The Generation of Malay Lexicon

Kalaiarasi Sonai Muthu Anbananthen, Surendran Selvaraju and Jaya Kumar Krishnan

Department of Information Science and Technology, Multimedia University, Melaka, Malaysia

Article history

Received: 22-08-2016

Revised:

Accepted:

Corresponding Author:

Kalaiarasi Sonai Muthu

Anbananthen

Department of Information

Science and Technology,

Multimedia University, Melaka,

Malaysia

Email: Kalaiarasi@mmu.edu.my

Abstract: The most important element in analyzing sentiment in text is to assign polarity to the opinion words. Polarity means the positive, negative or neutral state of the opinion words. They are many methods or ways in determining the polarity of an opinion words. One of the methods is using lexicon-based method. Lexicons are digital library of opinion words together with the polarity of the words. Basically, there are 3 methods in developing lexicon-based approach which is manual, dictionary-based and corpus-based. For Malay language there is no available sentiment lexicon and also very limited sources. Thus, in this study we present the automation lexicon generation for Malay language using the dictionary approach. The detail description of the automation lexicon generation for Malay language is discussed in this study.

Keywords: Sentiment Analysis, Sentiment Lexicon, Malay Lexicon, Automation Lexicon Generation

Introduction

In this new era the usage of social media is increasing. People are using social media such as Facebook, twitter and blog to interact and comment. This has generated gigantic amount of data. These data contain customer reviews or opinions about products and services. This information is essential for businesses to plan their marketing and product development. Also, customer's feedback on services or products will influence the decision of other customers. Statistic shows that Malaysia is one of the top 9 countries using social media in the world in 2015 (Mander, 2014). About 85% of Malaysian are using social media where 55% of these users are using Malay language to comment and give feedbacks (San et al., 2015). Malay language is not only used in Malaysia but also in Brunei, Singapore, Indonesia Philippines, central eastern Sumatra Riau islands and Thailand. The overall number of people using Malay in the world is around 270 million. There are a lot of reviews and comments being upload daily in Malay language but there is no available tools to analysis these texts. It is time consuming and tedious for any organization or people, to read through this online texts and capture the orientation of these texts. Therefore it is crucial to develop a sentiment mining tool to analysis the sentiment in Malay texts.

Sentiment mining is a type of natural language processing which track the mood of people on a product or service. It is a growing research area which mainly

focus on knowledge discovery and information retrieval from text using Natural Language Processing (NLP) techniques (Liu, 2012; Ganeshbhai and Shah, 2015; Jamaluddin, 2008). The task of sentiment analysis is to determine the polarity of the opinion in a given text. Polarity is the indicator of sentiments in the sentence, which can be positive, negative or neutral opinions. Polarity review can be divided into three levels, which are document level, sentence level and feature level (Liu, 2012; Sadegh et al., 2012). In document level, the polarity of the whole document will be classified into positive, negative or neutral. Whereas, in sentence level, the whole document is break down to sentences and the polarity each object in the sentence is classified as positive, negative or neutral. Both, document and sentence level, classified the polarity of the object discussed in the document or sentence as positive, negative or neutral, but it does not explore the details of object details discussed. It does not look into the feature or aspects about the particular object, which user are commenting about. On the other hand, feature level allowed more fine grain analysis and provides more detail information about the object that people are commenting (Tribhuvan et al., 2014). The sentences or documents will be classified based on feature or aspect. This approach is also known as aspect level sentiment classification.

In order to execute opinion mining task, the polarity of the word need to be determined. The polarity of the opinion word can be determined by two major approaches which is machine learning and lexicon approach

(Paltoglou, 2014) (Sidorov *et al.*, 2013). Machine learning approaches can be divided into two techniques which are supervised learning and unsupervised learning. Supervised technique need labeled training data and test data. The training data is used to train the classifier whereas the test data is used to validate the performance of the classifier used. In unsupervised technique, labeled training data is not required. On the other hand, lexicon-based approach is to develop lexicon which comprise list of words together with their polarity to classify the opinion words. Lexicon method can be divided to manual, corpus and dictionary based approach.

There are many lexicons developed for English but limited resources and tools available for Malay text. Alsaffar and Omar (2015) developed a Malay lexicon to analyst sentiment in Malay text by translating English WordNet to Malay language. Polarity are manually annotated to these translated words and also values are assigned to these words by Malay linguistic expert manually. The drawback of this technique is, an English word when translated into Malay language would have different meaning, which will affect the polarity. Furthermore, the developed lexicon is not automated.

The nearest language to Malay is Indonesia language. There are several lexicon developed for Indonesia language (Vania *et al.*, 2014) but this lexicon cannot be directly applied to Malay language. The morphology and the meaning of Indonesia words are different. For an instance the word “bisa” in Indonesia language means “can” but in Malay language it means “poison”. Therefore in this study we present an automation lexicon generation method for Malay language, which are useful in sentiment analysis tasks. Dictionary approach will be used to develop the Malay lexicon from scratch using annotated texts.

Work Related

In this session, different methods of developing a lexicon are reviewed and summarized as below.

Hu and Liu (2004) used seed words with known positive or negative orientation and grow it with searching of their synonym and antonym in WordNet (Fellbaum, 1988). The synonym words share same orientation of one another and antonym words are opposite orientation from one another. In WordNet, the words are arranged into bipolar clusters (Tsytsarau and Palpanas, 2012). This mean each word synset connected with head word synset. The lexicon generated would have two columns which are column contain adjective and one more column to store the words corresponding polarity which is positive or negative. Table 1 shows the lexicon that is generated by Hu and Liu (2004).

Kim and Hovy (2004), seed words is developed by handcraft and the polarity is assign to those words. The words with known polarity are listed as seed words.

Table 1. Lexicon generated by Hu and Liu (2004)

Adjective	Polarity
Big	Positive
Small	Negative
Beautiful	Positive

Table 2. Lexicon generated by Kim and Hovy (2004)

Abysmal: NEGATIVE [+ :0.3811][- :0.6188]
Adequate: POSITIVE [+ :0.9999] [- :0.0484e-11]
Afraid: NEGATIVE [+ :0.0212e-04][- :0.9999]

Table 3. Lexicon generated by Blair

Positive	Negative
Good_a (7.73)	Ugly_a (-5.88)
Swell_a (5.55)	Dull_a (-4.98)
Naughty_a (-5.48)	Tasteless_a (-4.38)
Intellectual_a (5.07)	Displace_v (-3.65)

The seed words polarity are arranged in two columns which are positive and negative. This seed word is expended by using the synset in WordNet. The synonym words of negative will be assigned as negative and the antonym words will be assigned as positive. The words which are synonym with positive word will have the same polarity. Words that are antonym to positive would have the opposite polarity which is negative. In order to initiate the developing the seed words, verbs and adjectives with known polarity are selected. It is learned that not all the synonyms and antonyms can be used, this is due to some words that have opposite polarity and some were neutral. This is where they came about with the idea of measuring the sentiment instead of just assigning words to positive, negative or neutral. The strength of the sentiment is measured by a “+” for positive sentiment and “-” for negative sentiment. The word silly has sentiment strength of -0.12. The higher the value of the positive sentiment the greater the positivity while the greater the value of the negative sentiment represent more negativity. Table 2 shows the output that is generated. The word abysmal is assigned as negative since it has higher value of negative compare to its positiveness. This method facilitates to determine the polarity of ambiguous words. This method is also enable the polarity to be assigned more accurate and precisely.

Similarly *et al.* (2014) use a Markov random walk model over relatedness graph which produce a sentiment estimate for a given word. To generate the relatedness graph it uses words synonym and hypernym that is found in WordNet. A word is classified as negative or positive by comparing its mean hitting time. The sentiment weight of the word is determined by taking the ratio between the mean of the hitting time.

Blair-Goldensohn *et al.* (2008) use a hybrid method to develop the lexicon which consists of two methods

Commented [U1]: Ref. not mention in the refs. list

known as general lexicon sentiment classifier and machine learning classifier. The hybrid of these two methods is believed to optimize the system parameters on a large data set. In ordered to construct the lexicon, small initial seed words with their sentiment orientation know. The small initial seed words then expanded by referring their link of synonym and antonym in WordNet. This method is similar with Hu and Liu (2004). In addition to this method, each word in the lexicon is weighted. The words are weighted with confidence measure which uses to represent the closeness of that particular word to positive or negative. The lexicon generated comprises of two columns which are positive and negative. Table 3 shows the resulted lexicon. Each word in the lexicon is appended with simplified part of speech tags which are adjective, adverb, noun and verb. The calculated weigh is also present for each word in bracket. From the positive column, the word Good (adjective) has a weigh of 7.73. From the negative column, the word Displace (adjective) has a weigh of -3.65.

Williams and Anand (2009) built the same structure of lexicon as proposed by Blair. Similarly, the words in the lexicon are appended with weigh or polarity strength. The lexicon generated consist of two columns, a column contain words and another column contains weight correspond to the word. Positive and negative words are placed in one column. Rao and Ravichandran (2009) also proposed the similar structure where the part of speech and weigh are append to the words.

Velikovich *et al.* (2010) proposed web-derived polarity lexicon. The proposed lexicon contains phrases with their corresponding polarity of positive, negative or neutral. A graph propagation technique is used to append sentiment strength to the phrases. Two sets of seed phrases are assumed as input. The positive seed set is denoted as P meanwhile the negative is denoted N. Figure 1 shows the propagation algorithm to build the graph. Figure 2 and 3 depicts the web lexicon for positive and negative phrases respectively.

Wilson and Wiebe (2005) used subjectivity lexicon in Phrase-Level Sentiment Mining. The lexicon used comprises of 8000 of subjective clues, this subjective clues are the phrases and words used to show polarity state. The words in the lexicon used are single word clues. The word that is least use in context is marked as weak subjective (weaksubj) whereas the word that is frequent used in context is marked as strong subjective (strongsubj). The list of the subjective clues is further expanded using thesaurus and dictionary. Each word or clue in the lexicon is tagged with their prior polarity which is positive, negative or neutral. Example of words that are tagged as neutral are see, look and feel and intensifiers also will be marked as neutral such as shortly. Those neutral words are included in the lexicon because this neutral word can be a good clue as well to detect a sentiment.

```

Input:       $G = (V, E), w_{ij} \in [0, 1],$ 
            $P, N, \gamma \in \mathbb{R}, T \in \mathbb{N}$ 
Output:      $\text{pol} \in \mathbb{R}^{|V|}$ 
Initialize:  $\text{pol}_i, \text{pol}_i^+, \text{pol}_i^- = 0$ , for all  $i$ 
            $\text{pol}_i^+ = 1.0$  for all  $v_i \in P$  and
            $\text{pol}_i^- = 1.0$  for all  $v_i \in N$ 

1.  set  $\alpha_{ii} = 1$ , and  $\alpha_{ij} = 0$  for all  $i \neq j$ 
2.  for  $v_i \in P$ 
3.     $F = \{v_i\}$ 
4.    for  $t : 1 \dots T$ 
5.      for  $(v_k, v_j) \in E$  such that  $v_k \in F$ 
6.         $\alpha_{ij} = \max\{\alpha_{ij}, \alpha_{ik} \cdot w_{kj}\}$ 
         $F = F \cup \{v_j\}$ 
7.  for  $v_j \in V$ 
8.     $\text{pol}_j^+ = \sum_{v_i \in P} \alpha_{ij}$ 
9.  Repeat steps 1-8 using  $N$  to compute  $\text{pol}^-$ 
10.  $\beta = \sum_i \text{pol}_i^+ / \sum_i \text{pol}_i^-$ 
11.  $\text{pol}_i = \text{pol}_i^+ - \beta \text{pol}_i^-$ , for all  $i$ 
12. if  $|\text{pol}_i| < \gamma$  then  $\text{pol}_i = 0.0$ , for all  $i$ 

```

Fig. 1. Graph propagation algorithm

POSITIVE PHRASES		
Typical	Multiword expressions	Spelling variations
cute	once in a life time	loveable
fabulous	state - of - the - art	niece
cuddly	fail - safe operation	niice

Fig. 2. Positive web lexicon

NEGATIVE PHRASES		
Typical	Multiword expressions	Vulgarity
dirty	run of the mill	fucking stupid
repulsive	out of touch	fucked up
crappy	over the hill	complete bullshit

Fig. 3. Negative web lexicon

Figure 4 above shows the subjectivity lexicon that is used by Wilson and Wiebe (2005). The clues in this lexicon are compiled from a number of sources. Some of the clues are obtained from manually developed resources. A large portion of the clues are collected from the work reported by Riloff and Wiebe (2003). From Fig. 4 each line contain one clues with it condition being stated. The type is to differentiate whether the clue is a strongsubj or weaksubj, len is to depict the length of the word, word1to resemble the token or stem of the clue and pos1 is the part of speech of the clue. Stemmed1 in the file is to let us know either word is stemmed or not and prior polarity is to mark the clue with it polarity which is positive, negative or neutral.

Commented [U2]: Figs. should be:

1. Text font size in (10)
2. Writing style should be Times New Roman
3. Text in figures should be normal not as bold

```
{type=weaksubj len=1 word1=abandoned pos1=adj stemmed1=n
priorpolarity=negative
type=weaksubj len=1 word1=abandonment pos1=noun stemmed1=n
priorpolarity=negative
type=weaksubj len=1 word1=abandon pos1=verb stemmed1=y
priorpolarity=negative
type=strongsubj len=1 word1=abase pos1=verb stemmed1=y
priorpolarity=negative}
```

Fig. 4. Subjectivity lexicon

@teniralcist	5	7	0
@brian_austin	5	7	0
@prayingmother	5	8	0
@drewlister	5	0	0
@durian_girl	5	8	0

Fig. 5. Unigrams pmilexicon.txt

compliment from	5	6	0
that's correct	5	6	0
at chicago	5	5	0
disney songs	5	5	0

Fig. 6. Bigrams pmilexicon.txt

```
p0007813-a neg nonabsorbent#1, nonabsorptive#1 -0.00036497131
00007990-a neg repellent#3, resistant#5 -0.000395879131
00013160-a pos concrete#2 0.000436737131
00013442-a pos objective#30, 000185065131
```

Fig. 7. Qwn-ppv lexicon

Mohammad *et al.* (2013) generated sentiment140 lexicon for sentiment analysis. Machine based approach is used. Pointwise Mutual Information (PMI) is used to generate the lexicon. From the score the sentiment orientation can be determined, positive sentiment would have a positive score whereas negative sentiment would have negative score. The corpus of sentiment140 is collected from Go *et al.* (2009) and is used to generate the lexicon.

In the unigram lexicon in Fig. 5, all the entries is single word, in bigram Fig. 6 all the entries would be two words while in pairs lexicon-the term would be pair such as unigram-unigram, unigram-bigram, bigram-unigram or bigram-bigram. All these pairs are obtained from large dataset of tweets. Score for the each entry would be real number of negative and positive.

San Vicente *et al.* (2014) develop lexicon using unsupervised dictionary based on method which is known as qwn-ppv (Q-WordNet as Personalized Page Ranking Vector). This qwn-ppv method is used for the automatic generation of the lexicon. Two approaches have been used in this method which is generating the lexicon from manually created seed word and generation of lexicon from set of seed word list created by Turney and Littman (2003). Figure 7 shows the lexicon

developed from Turney and Littman set of seed words. Each entry in the lexicon has an id to differentiate it from other entries. Part of speech of the word is also included with their short form-verb (v) and adjective (a). Each entry with positive sentiment will be append with positive (pos) and its positive weight.

It can be seen that most of the lexicons discussed above are developed for English. There is no lexicon available for Malay language. Besides, that there is less resource available on online and limited annotated text for Malay language. In order to analyze sentiment for Malay sentence there is a need to develop a Malay lexicon.

Lexicon Methodology

This Lexicon contains all the words with their respective polarity. Polarity is the indicator of sentiments in the sentence, which can be negative, positive or neutral opinions. Many general lexicons have been developed for English i.e., SentiWordNet but limited resources for Malay. A simple and an easy method to develop Malay lexicon is to translate the available English lexicon to Malay. However, this method is time consuming and also the “words” can have different meaning which can affect the polarity of the words. Translating polysemous words is difficult because it is unclear which meaning to use. Example the word light in the sentence “The phone is light”, the word light here will be translated to “lampu” when it supposedly mean “ringan”. The word “lampu” has the polarity of neutral whereas the word “ringan” brings positive polarity. Hence, translating available English lexicon to Malay will distorted the words polarity due to polysemous words. An important point in using lexicons is the domain from which the word comes from. Different domain may have different polarity for the same words. The word “fast” when it is used in medical domain such as “the cancer is spreading fast” its denote negative polarity while in telecommunication domain the usage of the opinion word “fast” brings the polarity of positive such as “The internet is quite fast at here”.

Basically there are 3 main approaches for developing lexicon which are manual approach, dictionary based approach and corpus based approach. The manual approach is tedious and time consuming. Dictionary based approaches use synonym and antonym concept of words. It automatically collect sentiment words based on manually created seed words. Corpus based approach on the other hand, employs patterns which can be found in set of sentences or documents. Based on the patterns, seed words are extended.

In this research we have used dictionary based approach to develop Malay lexicon. This is due to the factor that dictionary approach is usually more effective and contain all the words. Besides, dictionary approach is not domain specific which

Commented [U3]: Figs. should be:

1. Text font size in (10)
2. Writing style should be Times New Roman
3. Text in figures should be normal not as bold

mean it is applicable to all domains. First we have exploits a set of tweets in Malay. The Malay tweets was annotated manually with polarity labels. Malay lexicon (Mlex) was developed by bootstrapping words based on their synset relation with seed words. Synsets relation of a word are known as synonym and antonym of the word. In this research we have also developed Malay synset, a database contains the collection of synonym synset which are collected from various sources. The main source of comes from WordNet. As an example, the synonym for the word “kritikkan” are “komen” or “teguran”. Since these words are synonym, it is assume that they bring the same meaning. Hence same polarity. Based on this initial lexicon, we have extended and automated the process of lexicon generation.

Automated process of Lexicon

The preprocessed online sentence will be tagged with POS tag as shown as the sentence below. Based on the tagged sentence opinion words are identified. The words which are tagged as “KA” or “KK” are considered as the opinion words:

“[Unifi (KNK) punya (KT) kelajuan (KN) perlahan (KA)]”

After the opinion word is identified and extracted, the polarity of the word need to be determined. To determine the polarity of the opinion word, the word extracted will be checked in Mlex as shown in Fig. 8 flowchart and the pseudocode given in Fig. 9. If the opinion word exist in the Mlex the polarity of the particular word would be assigned. If the word is not found in Mlex, it is defined as unknown word. The synonym of the unknown word is searched through Malay Synset database. The unknown word may have one or more than one synonym words.

If the unknown word has one synonym word or if all the synonym words have same polarity, the polarity of the synonym word will be assigned directly to the unknown word. On the other hand, if the synonym of words have different polarity a weightage calculation will be performed to determine the polarity of the word.

Figure 9 shows the pseudocode for automated lexicon generation. Line 4, 5 and 6 shows the step in polarity search where polarity search function is called to determine the polarity of the opinion word. The size of initial Mlex and final Mlex should not be equal as add on of new opinion word would have taken place. In the polarity search function, line 5 and 6 where the synonym property of a word is used to get its polarity. Line 7 update the Mlex with the new word and its polarity.

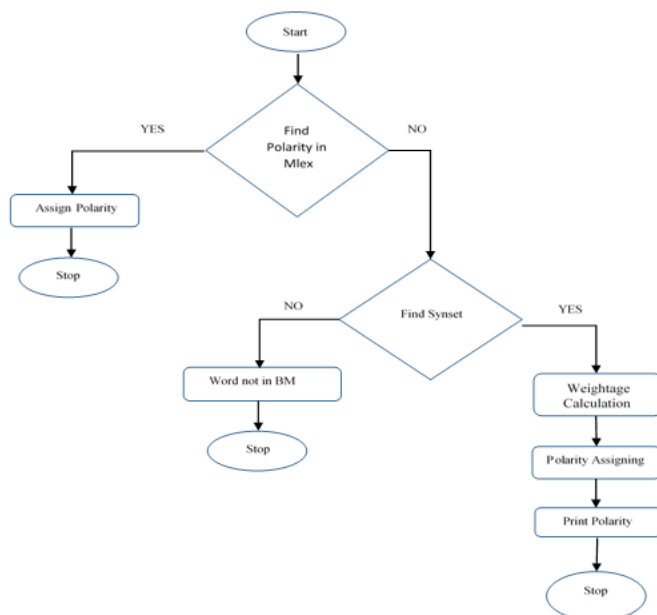


Fig. 8. Automated lexicon generation

■ ■ ■

For example the word laju". The synonym word of "Laju" is "cepat" and "merebak"; "cepat" has the positive polarity while "merebak" has the negative polarity. Hence, weightage calculation is used to determine the polarity of the unknown words. The calculation perform is by calculating the n of the synonym words to be positive, negative or neutral. The word which is not match to positive or negative polarity will be assigned as neutral for the polarity. The polarity which have the highest probability will be assigned to the opinion word. Equation 1 shows the formula used to calculate the weightage. W_i is the positive, negative or neutral words while \sum_w is the total words match. The procedure of assigning polarity is to determine the polarity of opinion word after the weightage calculation as shown in Fig. 10. After the polarity is assigned to the unknown word. The unknown word with its polarity would be added to the Mlex table:

$$p(w_i) = \log \frac{\sum w_i}{\sum_n w} \quad (1)$$

```

1 Step PolarityPrediction (opinion_word, Mlex)
2   Begin
3     Do {
4       Size x1 = of words in Mlex;
5       PolaritySearch (opinion_word, Mlex);
6       Size x2 = of words in Mlex;
7     } while (size x1 ≠ size x2);
8   End

1   PolaritySearch (opinion_word, Mlex)
2   Begin
3     For opinion_word wi
4       Begin
5         If (wi has synonym s in Mlex)
6           {wi polarity = s polarity;
7             Add wi with orientation to Mlex}
8         End for;
9   End

```

Fig. 9. Assigning the polarity orientation of opinion words

```

Begin
If (positive_weight & negative_weight = 0
AND neutral_weight > 0)
  Then wi = neutral;
If (positive_weight > neutral_weight)
  Then wi = positive;
If (positive_weight > negative_weight >
neutral_weight)
  Then wi = positive;
If (negative_weight > neutral_weight >
positive_weight)
  Then wi = negative;
End

```

Fig. 10. Polarity assigning

Experiment and Evaluation

This section evaluates the automated process of developing sentiment lexicons which are useful in sentiment analysis tasks. We carried out the experiments using the tweets of Telekom domain. About 9000 adjective words which were extracted from Malay tweets were used as the seed word to build Mlex. The polarities of these seed words were manually tagged. About 500 tweets were used to test the process of extracting and assigning polarity automatically. From the 500 tweets around 1330 opinion words which comprise of "KA" and "KK" are extracted. The 1330 opinion words are assigned with their respective polarity using the automation lexicon generation. The performance of the automatic lexicon generation is evaluated by using four indexes. The four indexes are accuracy, precision, recall and F1 score.

Table 4 shows the confusion matrix of polarity assigning for the 1330 opinion words. It can be seen that the true positive produced is 46% and the true negative produced is 39%. False negative and false positive are 2 and 11% respectively.

Table 5 shows the result produced by the automation lexicon generation. The recall and precision are distinguish good for the automatic word polarity assigning which are about 95 and 86%. In summary the Malay sentiment lexicon developed able to extract and automate the process with reasonable result. From evaluation 11% actual negative is predicted as positive due to ambiguous meaning word and there is no synonym entry for the particular word. The same reason goes for the 2% actual positive is predicted as negative. Based on the evaluation carried out, the performance of the Automation Malay lexicon is acceptable and able to increase the words in lexicon.

The work described above can be extended for phrases. From observation it is noted that opinion of a product not only depends on a single word but more than one words (phrase). Therefore our future work will be developing lexicon based on phrases.

Table 4. Results of automatic word polarity prediction

	Predicted positive	Predicted negative
Actual positive	46%	2%
Actual negative	11%	39%

Table 5. Result

	Automatic word polarity prediction (%)
Accuracy	86
Precision	81
Recall	95
F1	87

Conclusion

More and more Malaysians are buying products and share their reviews and feedbacks on online, thus a Malay lexicon is crucial to aid in sentiment retrieval. The opinion from this comments will benefit product manufacturers and consumers. Hence, in this study the automatic lexicon generation technique has been proposed and developed for Malay language. Since, there is no available resources for Malay language on online to be referred to, everything were done from scratch. Mlex and Malay synset database is also developed by collecting the possible synset of Malay words from other resources such as Malay dictionaries and Malay thesaurus which is quite tedious as Malay words are rich with synonym. The lexicon developed has been used to determine the polarity of Malay opinion words. Based on our experiment results, the performance of the proposed lexicon is acceptable.

Acknowledgement

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

Author's Contributions

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

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

JayaKumar Krishnan: Literature review.

Ethics

This article is unique and contains unpublished material. The corresponding author confirms that all of

Co-authors have read and approved the manuscript and there is no ethical issues involved.

References

- Alsaffar, A. and N. Omar, 2015. Integrating a Lexicon based approach and K nearest neighbour for Malay sentiment analysis. *J. Comput. Sci.*, 11: 639-639. DOI: 10.3844/jcssp.2015.639.644
- Blair-Goldensohn, S., K. Hannan, R. McDonald, T. Neylon and G.A. Reis *et al.*, 2008. Building a sentiment summarizer for local service reviews. Proceedings of the WWW Workshop on NLP in the Information Explosion Era, (IEE' 08).
- Vania, C., M. Ibrahim and M. Adriani, 2014. Sentiment lexicon generation for an under-resourced language. *Int. J. Comput. Linguist. Applic.*, 5: 59-72. <http://www.gelbukh.com/ijcla/2014-1/IJCLA-2014-1-pp-059-072-Sentiment.pdf>

- Fellbaum, C., 1988. WordNet: An Electronic Lexical Database. 1st Edn., MIT Press, Cambridge, ISBN-10: 026206197X, pp: 423.
- Ganeshbhai, S.Y. and B.K. Shah, 2015. Feature based opinion mining: A survey. Proceedings of the IEEE International Advance Computing Conference, Jun. 12-13, IEEE Xplore Press, pp: 919-923. DOI: 10.1109/IADCC.2015.7154839
- Go, A., R. Bhayani and L. Huang, 2009. Twitter sentiment classification using distant supervision. <http://www-cs.stanford.edu/people/alecmgo/papers/TwitterDistantSupervision09.pdf>
- Hassan, A., A. Abu-Jbara, W. Lu and D. Radev, 2014. A random walk-based model for identifying semantic orientation. *Comput. Linguist.*, 40: 539-562. DOI: 10.1162/COLI_a_00192
- Hu, M. and B. Liu, 2004. Mining and summarizing customer reviews. Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Aug. 22-25, Seattle, WA, USA, pp: 168-177. DOI: 10.1145/1014052.1014073
- Jamaluddin, N., 2008. Kriteria penggolongan kata: Bahasa Yunani, Inggris dan Melayu. *Jurnal Pengajian Umum Asia Tenggara*, 9: 139-162. http://www.ukm.my/jmalim/images/vol_09_2008/artikel%2011%20-%20norliza%20ms%20139-162.pdf
- Khan, K., B.B. Baharudin and A. Khan, 2009. Mining opinion from text documents: A survey. Proceedings of the 3rd IEEE International Conference on Digital Ecosystems and Technologies, Jun. 1-3, IEEE Xplore Press, pp: 217-222. DOI: 10.1109/DEST.2009.5276756
- Kim, S.M. and E. Hovy, 2004. Determining the sentiment of opinions. Proceedings of the 20th International Conference on Computational Linguistics, (CCL' 04), p: 1367-1367. DOI: 10.3115/1220355.1220555
- Liu, B., 2012. Sentiment analysis and opinion mining. *Synthesis Lectures Hum. Language Technol.*, 5: 1-167. DOI: 10.2200/S00416ED1V01Y201204HLT016
- Mander, J., 2014. GWI social summary. Global Web Index. <http://insight.globalwebindex.net/social>
- Mohammad, S.M., S. Kiritchenko and X. Zhu, 2013. NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets. Proceedings of the 2nd Joint Conference on Lexical and Computational Semantics, (LCS' 13), Atlanta, USA, pp: 321-327
- Turney, P. and M. Littman, 2003. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Trans. Inform. Syst.*, 21: 315-346. DOI: 10.1145/944012.944013

Commented [U4]: Ref. not discussed in the text

Commented [U5]: Ref. not discussed in the text

- Paltoglou, G., 2014. Sentiment analysis in social media. In: Online Collective Action: Dynamics of the Crowd in Social Media, Agarwal, N., M. Lim and R.T. Wigand (Eds.), Springer, Wien, ISBN-10: 3709113407, pp: 3-17.
- Rao, D. and D. Ravichandran, 2009. Semi-supervised polarity lexicon induction. Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, (ACL' 09), ACM, pp: 675-682. DOI: 10.3115/1609067.1609142
- Riloff, E. and J. Wiebe, 2003. Learning extraction patterns for subjective expressions. Proceedings of the Conference on Empirical Methods in Natural Language Processing, (NLP' 03), ACM, pp: 105-112. DOI: 10.3115/1119355.1119369
- Sadegh, M., R. Ibrahim and Z.A. Othman, 2012. Opinion mining and sentiment analysis: A survey. Int. J. Comput. Technol., 2: 171-178.
- San Vicente, I., R. Agerri, G. Rigau and D.S. Sebastián, 2014. Simple, Robust and (almost) unsupervised generation of polarity lexicons for multiple languages. <http://www.aclweb.org/anthology/E14-1#page=114>
- San, L.Y., A. Omar and R. Thurasamy, 2015. Online purchase: A study of generation Y in Malaysia. Int. J. Business Manage., 10: 1-7. DOI: 10.5539/ijbm.v10n6p298
- Sidorov, G., S. Miranda-Jiménez, F. Viveros-Jiménez, A. Gelbukh and N. Castro-Sánchez *et al.*, 2013. Empirical Study of Machine Learning Based Approach for Opinion Mining in Tweets. In: Advances in Artificial Intelligence, Batyrshin, I. and M.G. Mendoza (Eds.), Springer, Berlin, ISBN-10: 3642378072, pp: 1-14.
- Tribhuvan, P.P., S.G. Bhirud and A.P. Tribhuvan, 2014. A peer review of feature based opinion mining and summarization. Int. J. Comput. Sci. Inform. Technol., 5: 247-250.
- Tsytssarau, M. and T. Palpanas, 2012. Survey on mining subjective data on the web. Data Min. Knowledge Discovery, 24: 478-514. DOI: 10.1007/s10618-011-0238-6
- Velikovich, L., S. Blair-Goldensohn, K. Hannan and R. McDonald, 2010. The viability of web-derived polarity lexicons. Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics, (ACL' 10), ACM, pp: 777-785. <http://dl.acm.org/citation.cfm?id=1858118>
- Vinodhini, G. and R.M. Chandrasekaran, 2012. Sentiment analysis and opinion mining: A survey. Int. J.
- Williams, G.K. and S.S. Anand, 2009. Predicting the polarity strength of adjectives using wordnet. Proceedings of the 3rd International AAAI Conference on Weblogs and Social Media, (WSM' 09), pp: 346-349. <https://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/214>
- Wilson, T., J. Wiebe and P. Hoffmann, 2005. Recognizing contextual polarity in phrase-level sentiment analysis. Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, (NLP' 05), pp: 347-354. DOI: 10.3115/1220575.1220619

Commented [U6]: Ref. not discussed in the text