Sentiment Classification of Malay Newspaper Using Immune Network (SCIN)

Norulhidayah Isa, Mazidah Puteh, and Raja Mohamad Hafiz Raja Kamarudin

Abstract— **The advancement of internet technology and machine learning techniques in information retrieval make Sentiment mining or analysis become popular among researchers. There are many sentiment analysis researches that have been done on English text using various machine learning techniques. However, there are very limited researches on Malay text sentiment analysis. This research focuses on preprocessing techniques for stemming Malay text namely Reverse Porter Algorithm (RPA) and Backward-Forward Algorithm (BFA) and Artificial Immune Network (AIN) for extracting sentiment from Malay newspaper articles. Data representation is also important where the data must be converted into suitable form for Artificial Immune Network Algorithm to work. To represent the data, vector space representation is used with three parameters represent the actual word, the frequency of occurrence of the word in a particular sentence and the polarity of the sentences. Lastly, the sentiment analysis algorithm was developed using immune network from Artificial Immune System (AIS). The result from stemming the Malay text using new reverse Porter algorithm shows some improvement in processing time compared to backward-forward algorithm. However, the sentiment analysis accuracy using AIN with both stemming techniques show almost similar result. In the future, thorough study on artificial immune system techniques and comparative study on other machine learning techniques for sentiment analysis is required for better result.**

***Index Terms*— Artificial immune system, Artificial Immune Network, text mining, sentiment analysis, sentiment mining, Malay text, stemming technique.**

I. INTRODUCTION

Sentiment Analysis or Opinion Mining is a study that attempts to identify and analyze emotions and subjective information from text. Since early 2001, the advancement of internet technology and machine learning techniques in information retrieval make Sentiment Analysis becomes popular among researchers[1]. Besides, the emergent of social networking and blogs as a communication medium also contributes to the development of research in this area.

Most of the researchers analyze sentiments in subjective text such as blogs and social networking[2]. Subjective text type contains a direct message to their user. The writers also have a specific target in their mind when they are writing the articles or comments. The writer expresses their opinions freely and the reader knows the message is either negative or positive. It is different compared to news article. The target readers for news article are general and consist of various groups of people. Besides, the journalists try to contain themselves from expressing direct opinion especially when it involves a sensitive issue. They may end up expressing their opinion through a complex argument structure, or quoting others in expressing their own feeling. In Sentiment Analysis research, news article has gained less attention compared to subjective data type [3].

News articles undeniably give a big impact in our everyday live. [4] Revealed in their papers is the impact of financial news towards economy where financial market is very sensitive towards financial news. Furthermore, measuring sentiment accumulated in stock market news articles - also help us in stock market prediction [5]. Besides, government can use it in preventing negative viral effect especially when it involves critical issues.

This study focuses on Malay stemming algorithms namely Reverse Porter Algorithm (RPA) and Backward-Forward Algorithm (BFA) and Immune Network Algorithm (INA) for extracting sentiment from Malay newspaper articles. This is because Malay language differs in term of structure and complexity from the English language.

The organization of the paper; Section II explains the related work of sentiment mining, stemming algorithms and artificial immune system algorithms. Section III elaborates - the methodologies carried out in the research. Section IV discusses the findings and Section V concludes the research and states the future suggestions.

II. RELATED WORKS

A. *Sentiment Analysis*

Sentiment analysis or mining refers to the application of Natural Language Processing, Computational Linguistics, and Text Analytics to identify and extract subjective information in source materials [6]. Sentiment mining extracts attitude of a writer in a document includes writer's judgement and evaluation towards the discussed issue. Sentiment analysis allows us to identify the emotional state

This research is supported in part by Fundamental Research Grant Scheme (FRGS) under the ninth Malaysian Plan (RMK-9), Ministry of Higher Education (MOHE) Malaysia. The grant no is 600-RMI/ST/FRGS 5/3/Fst (208/2010).

Norulhidayah Isa is with Universiti Teknologi MARA (UiTM) . She is now in Computer Science Department, UiTM Terengganu, 23000 Dungun, Terengganu, Malaysia. (e-mail: norul955@tganu.uitm.edu.my).

Mazidah Puteh is with Universiti Teknologi MARA (UiTM). She is now in Computer Science Department, UiTM Terengganu, 23000 Dungun, Terengganu, Malaysia. (phone: 6019-9316631; fax: 609-8400288; e-mail: mazidahputeh@tganu.uitm.edu.my).

Raja Muhammad Hafiz Raja Kamarudin is with Universiti Teknologi MARA. She is now in Computer Science Department, UiTM Terengganu, 23000 Dungun, Terengganu, Malaysia (e-mail: hafizmml89@gmail.com).

of the writer during writing, and the intended emotional effect that the author wishes to give to the reader. In recent years, sentiment analysis becomes a hotspot in numerous research fields, including natural language processing (NLP), data mining (DM) and information retrieval (IR). This is due to the increasing of subjective texts appearing on the internet [7]

In previous research, Machine Learning is commonly used to classify sentiment from text. This technique involves with statistical model such ad Support Vector Machine (SVM) and Näive Bayes (NB) [8]. For example, [5] the applied Contextual Entropy Model to classify sentiment in stock market news.

Besides, there are several attempts by researcher to classify sentiment using Biological Inspired Computational Algorithm. [9](dear writer, this part is rather incomplete and hanging, could you justify) developed entropy weighted genetic algorithm (EWGA) where the researcher hybridized genetic algorithm with information gain heuristic for feature selection to classify sentiment in web forum content for English and Arabic languages.

B. Preprocessing

Preprocessing is a process to prepare the data for mining process. It has been proven that text preprocessing can speed up the mining process. [10] implemented two methods in preprocessing Malay text which were stop word remover and Malay stemming processes.

Stop word is defined as word that has no significant sentimental on the particular domain. In classifying sentiment value from text, any non-sentiment words must be removed for better algorithm performance [11].

Stemming is the process of reducing inflected or derived words to their stem or root word. Porter Algorithm that has been invented by Martin F. Porter in 1980 is one of the well known Stemming Algorithm.

The algorithm is based on the idea that affixes in English are mostly made-up of a combination of smaller and simpler suffixes. The stripping process is performed on a series of steps, which simulates the inflectional and derivational process of a word. At each step, a certain suffix is removed by means of substitution rules. A substitution rule is applied when a set of conditions or constraints attached to this rule holds.

The usage of affixes in English language and other European languages is less complex compared to Malay [12] and Arabic [13] language. It is because for English language, stemmers only concern with the removal of the suffixes.

Affix is the verbal element that is attached at the beginning of the word (prefix), at the end of the word (suffix) or at the middle of the word (infix). In Malay language, more than one affixes can be attached to a word at the same time. It may contain both affixes, a prefix and a suffix, and this is also known as prefix-suffix pair [12]. As an example, let's observe the word 'pemakanan' (diet). The root word is 'makan' (eat), and then the prefix 'pe' is added at the beginning and the suffix 'an' is added at the end of the word. Some words have more than one suffix or prefix. Such as 'makanannya' (his/her food) which has two suffixes, 'an' and 'nya'.

Affix complexity in Malay Languages causes Porter Algorithm not suitable to stem the language. The problems with porter algorithm are; 1) Over-stemming: The algorithm over stems the word resulting wrong root word formed. Example, "memakan" is over stemmed to "mak", while the right word is "makan". 2) Under-stemming: A situation when the algorithm has found a root word even though the stemming process is not enough to produce the right word. Example, "memakannya" is under stemmed to "memakan", while the real result is "makan". 3) Unchanged: The algorithm cannot stem the word because there are not enough rules to use. Example, "pelajari" cannot be stemmed without "i" suffix rules. 4) Spelling exception: The algorithm cannot find the matching root words when the stemmed words do not exist in the library of root words. Example, "menyayangi" is stemmed to "yayang" which is not a root word. The real word is "sayang".

[14] proposed Backward-Forward Stemming Algorithm for Malay text to overcome those problem. Backward-forward changes the above steps by removing suffixes first and followed by removing prefixes and prefixes + suffixes later. The author claims that by doing this, it has improved the efficiency of the algorithm. Details about this stemming algorithm will be discussed in Section III: Methodology.

C. Artificial Immune System

Artificial Immune System (AIS) is a Biological Inspired Computation techniques inspired by the biological immune system [15]. Artificial Immune System methods mainly use three immunological principles, including the Immune Network Theory, the mechanisms of Negative Selection and the Clonal Selection Principle.

Several applications have been developed using Immune Network algorithm. [16] and [17] apply Artificial Immune Network in classifying heterogeneous data, while [8] use AIS algorithm to classify sentiment in Malay Movie reviews.

III. METHODOLOGY

A. Data Collection

As we discussed before, the most commonly used in sentiment mining were taken from blog, twitter and web review which focusing on sentences that expressed sentiment directly as explained in [9],[18],[19], and [20]. One of the challenges in this research area is to develop a sentiment mining model that can process a multilingual language [21]. Inspired by this challenge, this research will focus on formal Malay language from Malaysian newspaper that is Berita Harian.

The researcher chooses Berita Harian because it is a national newspaper in Malaysia. It also has a strict authoring policy, thus, grammar and spelling mistakes are very unlikely to happen. Furthermore, the language used in the newspaper is very formal and compels to the Malay language standard. The data were collected from July to December 2011. It consists of 1080 of sentences available with an average of 40 words per sentence totaling to an estimation of 4000 words available for processing. The

ISBN: 978-988-19252-9-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Malay Language Domain expert is hired to verify the data and identify its sentiment polarity whether the sentence is positive or negative. Positive valued sentence means that the sentence has positive value in respect to its domain. For example, a sentence about recovery of a recent natural disaster is a positive sentence while a sentence that describes a falling stock in the market seems to be negative value for economy.

B. Text Preprocessing

For this research, three common steps in text preprocessing have been implemented which are Stemming, Stop Word Removal and Word Tokenizer.

Stop Word Remover

Stop word removal is the process of removing stop words from the query to ignore common words known as stop word and return more relevant results in the end of the process. In easy word, stop word removal is the process of removing the words that have high frequency which are not important to the meaning of the sentence. Words such as; ‘a', ‘the', ‘or' are likely to be considered as stop words which have been listed in [22]. For this research, stop word suggested by [10] has been used.

Stemming

For this research, the researcher has invented a Reverse Porter Algorithm based on Porter Algorithm and compares it with Backward-Forward Algorithm for the stemming process.

Backward-Forward Algorithm

The Backward-Forward algorithm consists of two main processes, backward stemming and forward stemming as illustrated in the left side of the diagram. In the backward stemming, there are six sub-processes;

- Dual-word conversion
  - o Dual-word conversion is a process to remove Malay dual word into one single word only.
  - o For example, the word “berlari-lari” will be reduced into “berlari”.
- Check against the root dictionary
  - o Root dictionary is the collection of all root words obtained from Kamus Dewan 2005 edition.
  - o The word will be first compared against the root dictionary. If there is a match in the root dictionary, the engine will output the word bypassing all the subsequence process.
- Remove suffix
  - o In this phase, the engine will remove the suffix from the word. The operation is by finding any sub word (a part of word from input word) that matches any suffixes (the list of suffixes gathered from a Malay grammar textbook).
  - o The algorithm works by creating a sub word using suffixes as the base. For example, suffix,”nya” is 3 characters long, the engine will get the last 3 character from the original word and matches it with the suffixes. If it matches the suffix, the sub word will be removed from the word thus producing a new word, “akhir”.

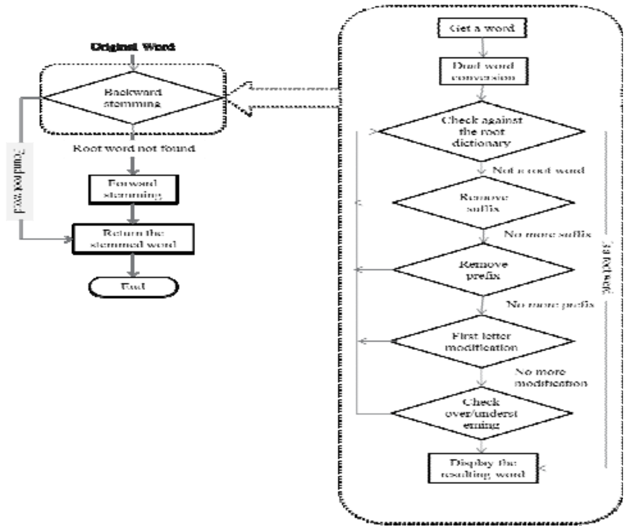

Fig. 1 Backward-Forward Algorithm [14]

  - o Lastly, the resulting word will be compared against the root dictionary for confirmation.
- Remove prefix
  - o In this phase, the engine will remove the prefix from the word. The process is similar to the process above but the engine will get the first 3 characters (or any number of characters that is the same with the prefix) and the sub word will be matched with the prefix.
  - o For example, the prefix, “ber” that has 3 characters length will be used as the base. The engine will get the first 3 characters from the given word, for example, “bekerja”, and the result will be “bek”. Both prefix and sub word will be matched against each other. If they matched, the sub word will be removed from the given word. In this case, they are not matched.
  - o Lastly, the resulting word will be compared against the root dictionary for confirmation.
- First letter modification
  - o In Malay language, adding prefixes to some words will transform a part of the word, by not simply adding the prefix in front of the word.
  - o For example, the word “ketuk” when combined with prefix “me” will become “mengetuk”. Notice that the “K” character has transformed into “NG” characters. This has been discussed in chapter 2.
  - o In this phase, the engine will transform those special words back to their original words after removing the prefixes.
  - o Lastly, the resulting word will be compared against the root dictionary for confirmation.
- Check over or understemming
  - o In this phase, the engine will check the resulting word for over or understemming problems (this has been discussed in chapter 2).
  - o If any, the engine will try to repair the word. If the engine cannot repair the word, it is deemed as failure.
- Lastly, the resulting word will be compared against the root dictionary for confirmation.

However, even with the substantial increment in efficiency, the Backward-Forward algorithm still cannot fully solve the four problems in stemming, namely, over or

ISBN: 978-988-19252-9-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

under stemming, spelling mistake and unchanged. Because of this, we invented reverse Porter Algorithm.

Reverse Porter Algorithm

The basic idea of this algorithm is to reverse the whole process of Porter algorithm in order to get the result. The main concept of Porter Algorithm is to use a reduction technique where the given words will be reduced to its root form, which can be presented as:

$$w - w' = rW$$

Where w is word, w' is suffixes, and rW is root word. Contradict to Porter algorithm; the Reverse Porter algorithm will make some repetitions to all the root words that contain in root dictionary and do some combinations with all suffixes available to produce the exact same words as the input word. If the resulting words and the input words matched, then we have the root words and the suffix used to create the given words.

This can be presented as:

$$ArW + Aw' = rS; rS = gW$$

In this representation, ArW is all root words, Aw' is all suffixes, rS are the result Word and gW is the given word. In logic, this operation is hard to do in computational, as it has to repeat all root words and suffixes. However, for each achievement, the resulting words will be stored in a library so that the engine will not have to repeat all over if it encounters the same word, hence reduces the computational times greatly. The Reverse Porter Algorithm is shown in Figure 2.

Word Tokenizer

Word Tokenizer algorithm is a process to discover the occurrence of each word in a sentence. It is a simple process that needs some identifying processes of any repeating words and then reduces it to one word with the number of occurrences.

After all the 3 processes have been executed, an XML file will be output by the preprocessing engine storing the processed data. The output after the preprocessing stage, will be divided into two categories; training and testing with 80:20 ratio. The deviation is random.

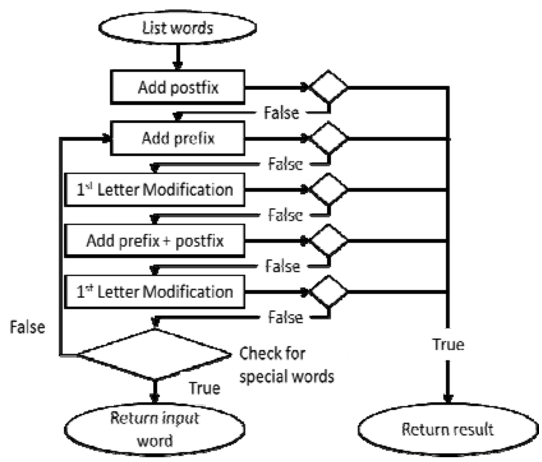

Fig. 2 Reverse Porter Algorithm

Data Representation

Data representation is a process of converting the data into suitable form for artificial immune network algorithm. To represent the data, we use vector space representation. Each word will be converted according to following format

Data = {word, frequency, domain, polarity}

Fig. 3 Word representation in vector space

Word1 = {"makan",4,"politik","positive"}

Fig. 4 Example of word representation using Vectors Space

The first parameter represents the actual word. The second parameter represents the frequency of occurrence of the word in a particular sentence. Third parameter represents the domain of the word and the last parameter represents the polarity of the sentences. Then, those words are combined in one sentence as shown in the figure five below. For example, sentence "Saya[2] makan[1] nasi[2] sedap[1] suka[1] itu[1]" which has been preprocessed from the original sentence, "Saya makan nasi sedap". Saya suka nasi itu" will be converted in this form of representation

S1[Word1{saya,2,politik,positive}, Word2 {makan,1,politik,positive}, Word3{nasi,2,politik,positive }, Word4{sedap,1,politik,positive }, Word5 {suka,1,politik,positive }, Word6{itu,1,politik,positive }]

Fig. 5 Example of data representation in sentences. S1 means sentence one and Word1 means the first word in the sentence

C. *Artificial Immune Network Algorithm*

Next phase of this research is Algorithm development. During this phase, the algorithm for sentiment mining is developed. The algorithm was developed based on the aiNet algorithm proposed by [23].

Artificial Immune System (AIS) is a class of Biological Inspired Computation techniques that inspired by the principles and processes of the vertebrate immune system. The vertebrate immune system is a structure of diverse sets of cells and molecules that work together with other systems to maintain homeostatic state. The primary function of the immune system is to protect human bodies from infectious agents such as viruses, bacteria and other parasites which are also known as pathogens. AIS consist of several techniques which are Clonal Selection Algorithm, Negative Selection Algorithm and Artificial Immune Network Algorithm.

The training algorithm is based on Artificial Immune Network algorithm. However, slight alterations have to be made as the original aiNet algorithm does not process text-type input.

The engine begins with iterating each word (antigen). Each antigen will be compared against the library (antibodies-AB) in the memory cell in affinity measurement process. Affinity measurement process is a process of measuring the difference between antibodies and the antigens and selecting the best antibodies that match the antigen. The formula of affinity measurement is described below;

$$\text{Antibody } (w, d, f, p) - \text{antigen } (w, d, f, p) < a$$
$$a = \text{accepted rate (user defined variable)}$$

First, the engine will find antibodies with the same word and polarity. If there are any, the engine will iterate all of the antibodies and find the nearest antibody by finding the smallest acceptable value of frequency of occurrence that falls under the accepted rate which is user defined variable. However, if there is no antibody reacts to the antigen, the antigen will then be added to the memory cell as the new antibody.

The best antibody that has been selected will then go through cloning process. The rate of cloning is proportionate to the affinity value.

All the clones will then be mutated in the following mutation process. Mutation of the clones is done by adjusting the frequency of the word. The formula of mutation is described below:

newFrequency = oldFrequency + (rand() - 0.5)*mutationRate
mutationRate is user defined.

*Rand() function will return a floating point between 0 and 1 inclusive.

By deducting the value of rand() with 0.5, we can get a floating point between -0.5 to 0.5 inclusive. Then, the value will be multiplied by mutation rate variable.

The reason why the rand() function must be deducted with 0.5 is to produce a mutated antibody with lower frequency than the original as not all antigens have greater frequency than the antibodies.

After that, all the mutated clones are compared to the antigen with the same affinity measurement process. The best mutated clone will be picked and added into the memory cell.

All the above process will be repeated until all antigens have been processed.

In the network suppression process, all the antibodies will be compared against each other. If there are many antibodies with similar or near frequency, those antibodies will be combined to become a single antibody. Notice that with a rate of 0, the antibodies will not be suppressed.

The suppression is done by comparing antibodies against each other. Antibodies that are similar in word, polarity and domain parameter will be grouped together. Then the distance between them is calculated by deducting the frequency of those antibodies. If the value falls below the suppression rate, the antibodies will be destroyed and replaced by a single antibody where its frequency is the median of the destroyed antibodies.

For example,

Antibody1("saya",1,"politik","positive")
Antibody2("saya",2,"politik","positive")

will be suppressed if the suppression rate is 1 or higher, resulting to a new antibody;

newAntibody("saya",1.5,"politik","positive")

After the development phase has finished, training phase is carried out. The processed text data are presented to the algorithm and a library of negative and positive antibody will be produced.

The algorithm used can be summarized as below.

Initialization: Initialized an antigen (**Ag**) by assigning data to Ag. Ag is agroup according to the sentences. For example S_1**[**$\mathbf{Ag_1, Ag_2, Ag_3 \ldots Ag_n}$**]**, S_2**[**$\mathbf{Ag_1, Ag_2, Ag_3 \ldots Ag_n}$] … $\mathbf{S_n}$**[**$\mathbf{Ag_1, Ag_2, Ag_3 \ldots Ag_n}$**]**.
During the initialization process, Memory Cell (**MC**) is set to null.
Repeat for each $\mathbf{S_i}$

- For Each $\mathbf{Ag_i}$ do:
  - *Clonal Selection and Expansion*: Determine affinity by presenting **Ag** to **ABs** in the memory cell**.** If there is no **AB** reacted to the **Ag,** the **Ag** will then be added to memory cell as new **AB.** Select the **AB** with highest affinity value and reproduce (clone). Clonal rate defined by user.
  - *Affinity Maturation:* Mutate the clones by adjusting the frequency. Measure the affinity of the clones and select clone with the highest affinity value to be placed in MC.
  - *Network Construction:* incorporate AB

Network Suppression: Merge the **AB** with similar frequency with others.

D. Classification Process

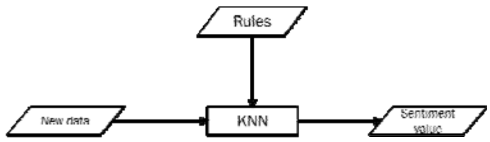

Fig. 6 Classification Process (new data: antigen; rules: antibody)

Figure six illustrates the classification process. In short, k^{th} Nearest Neighbour (KNN) will be used in order to extract the sentiment value of the sentences based on the rules gathered by the AIN algorithm.

Sentence is presented to the sentiment library generated during the training process. Then, kNN is used to count how many positive and negative antibodies in total of k value that respond to the antigen. If the positive are more than the negative antibodies, the sentiment value for that particular sentence is positive or vice-versa.

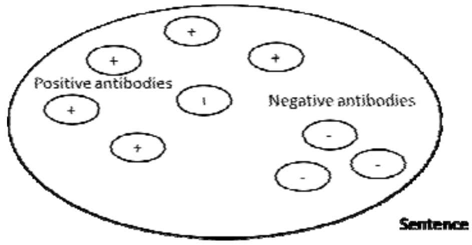

Fig. 7 Classification Process

Figure seven shows that there are more positive than negative antibodies. Then the classifier will classify the sentence as a positive-valued sentence.

IV. Result and discussion

This research operates on 1080 sentences from newspapers that are splitted into 80% training dataset and 20% testing dataset. Result for preprocessing using Reverse Porter Algorithm and Backward-Forward Algorithm for stemming the words is shown in Table 1.

TABLE I
STEMMING TECHNIQUE COMPARISONS

	1st run		2nd run	
	Accuracy	Time	Accuracy	Time
Reverse Porter	96	595	96	44
BackwardForward	90	322	90	316

TABLE II
SCIN ACCURACY COMPARISONS

Stemming Algorithm	Accuracy
Reverse Porter	53.67
BackwardForward	53.07

Table II shows the accuracy performance for the sentiment analysis on Malaysian Newspaper using Artificial Immune Network. Accuracy for SCIN using Reverse Porter Algorithm for stemming is 53.67% and Backward-forward is 53.07% with the difference of 0.60% only.

V. Conclusion

The main idea of this research is to introduce the new technique for stemming of Malay text called Reverse Porter Algorithm and to compare it with the existing technique called Backward-forward algorithm. The research also examines the Immune Network Algorithm to harvest sentiment values from Malay text. As the research continues, a better understanding on how sentiment can be harvested from Malay text is gained. This research tries to explore new ways to harvest sentiment values from text by utilizing bio-inspired machine learning technique called Immune Network algorithm.

Further researches are required to increase the efficiencies of reverse porter algorithm and immune network algorithm in sentiment analysis for Malay text. Hybrid techniques can be considered to upgrade the accuracy of Malay text sentiment analysis.

References

[1] B. Pang and L. Lee, *Opinion mining and sentiment analysis*: Now Pub, 2008.
[2] A. Balahur and R. Steinberger, "Rethinking Sentiment Analysis in the News: from Theory to Practice and back," *Proceeding of WOMSA*, 2009.
[3] A. Balahur, R. Steinberger, M. Kabadjov, V. Zavarella, E. Van Der Goot, M. Halkia, B. Pouliquen, and J. Belyaeva, "Sentiment analysis in the news," in *Proceedings of LREC*, 2010.
[4] J. Kleinnijenhuis, F. Schultz, D. Oegema, and W. van Atteveldt, "Financial news and market panics in the age of high-frequency sentiment trading algorithms," *Journalism*, 2013.
[5] L.-C. Yu, J.-L. Wu, P.-C. Chang, and H.-S. Chu, "Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news," *Knowledge-Based Systems*, vol. 41, pp. 89-97, 2013.
[6] W. Wang, "Sentiment analysis of online product reviews with Semi-supervised topic sentiment mixture model," in *Fuzzy Systems and Knowledge Discovery (FSKD), 2010 Seventh International Conference on*, 2010, pp. 2385-2389.
[7] B. Liu, "Sentiment analysis: A multi-faceted problem," *the IEEE Intelligent Systems*, 2010.
[8] N. Samsudin, M. Puteh, A. R. Hamdan, and M. Z. A. Nazri, "Is artificial immune system suitable for opinion mining?," in *Data Mining and Optimization (DMO), 2012 4th Conference on*, 2012, pp. 131-136.
[9] A. Abbasi, H. Chen, and A. Salem, "Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums," *ACM Transactions on Information Systems (TOIS)*, vol. 26, p. 12, 2008.
[10] A. Kwee, F. Tsai, and W. Tang, "Sentence-level novelty detection in English and Malay," *Advances in Knowledge Discovery and Data Mining*, pp. 40-51, 2009.
[11] E. Dragut, F. Fang, P. Sistla, C. Yu, and W. Meng, "Stop word and related problems in web interface integration," *Proceedings of the VLDB Endowment*, vol. 2, pp. 349-360, 2009.
[12] F. Ahmad, M. Yusoff, and T. M. Sembok, "Experiments with a stemming algorithm for Malay words," *Journal of the American Society for Information Science*, vol. 47, pp. 909-918, 1996.
[13] S. Khoja and R. Garside, "Stemming Arabic Text," *Lancaster, UK, Computing Department, Lancaster University*, 1999.
[14] A. SAMSUDIN, "Teknik Stemming Bahasa Melayu Berasaskan Algoritma Porter," Universiti Teknologi Malaysia, 2003.
[15] L. N. De Castro and F. J. Von Zuben, "aiNet: An Artificial Immune Network for Data Analysis," 2001.
[16] M. Puteh, A.R. Hamdan, K.Omar dan A.Abu Bakar, " Flexible Immune Network for Mining Heterogeneous Data," *LNCS 5132, hlm 232-241 @ Springer-Verlag Berlin Heidelberg*, 2008.
[17] M. Puteh, A.R. Hamdan, K.Omar dan M.T.H. Mohd Tajuddin, "Artificial Immune Network: Classification on Heterogeneous Data," *Machine Learning, InTech, ISBN: 978-953-307-033-9*, 2010.
[18] W. Claster, M. Cooper, and P. Sallis, "Thailand--Tourism and Conflict: Modeling Sentiment from Twitter Tweets Using Naïve Bayes and Unsupervised Artificial Neural Nets," 2010, pp. 89-94.
[19] Y. Liu, X. Huang, A. An, and X. Yu, "ARSA: a sentiment-aware model for predicting sales performance using blogs," presented at the Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, Amsterdam, The Netherlands, 2007.
[20] V. Sehgal and C. Song, "SOPS: Stock Prediction Using Web Sentiment," presented at the Proceedings of the Seventh IEEE International Conference on Data Mining Workshops, 2007.
[21] H. Chen and D. Zimbra, "AI and Opinion Mining," *IEEE Intelligent Systems*, vol. 25, pp. 74-80, 2010.
[22] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to information retrieval* vol. 1: Cambridge University Press Cambridge, 2008.
[23] N. Leandro, D. CASTRO, J. FERNANDO, and V. ZUBEN, "Immune and neural network models: theoretical and empirical comparisons," *International Journal of Computational Intelligence and Applications*, vol. 1, pp. 239-257, 2001.