Towards Sentiment Analysis Application in Housing Projects

Nurul Husna Mahadzir[1], Mohd Faizal Omar[1], Mohd Nasrun Mohd Nawi[2]
[1]School of Quantitative Sciences,
Universiti Utara Malaysia, 06100 Sintok, Kedah.
[2]School of Technology Management and Logistic,
Universiti Utara Malaysia, 06100 Sintok, Kedah, Malaysia.
nhusna84@gmail.com

Abstract—In becoming a develop nation by 2020, Malaysia Government realized the need in providing affordable house to the public. Since Second Malaysia Plan, government has implemented various affordable housing projects and it continues until recent Malaysia Plan. To measure the effectiveness of the initiatives taken, public opinion is necessary. A social media platform has been seen as the most effective mechanism to get information on people's thought and feeling towards certain issues. One of the best ways to extract emotions and thoughts from what people post in social media is through Sentiment Analysis (SA). This paper will propose a new framework that focuses on the application of sentiment analysis to assist the decision maker in understanding the real voice of the public in regard to property industry in Malaysia. The framework will consist of two components; sentiment classification at feature/aspect level and sentiment visualization to show the results of the analysis.

Index Terms—Housing Projects; Sentiment Analysis; Text Mining.

I. INTRODUCTION

Housing prices always stand as major issue for most Malaysians. It is considered as no longer affordable especially for the low and middle income household and the concern leads to legal issue, planning policy, economic development and industrial organization [1]. Be in the news or social media, public are raising concerns about higher house prices and mortgages, and the increased expenses associated with owning a house. Due to that matter, government has implemented various affordable housing programmes since Tenth Malaysia Plan and the programme is continued and improved during Eleventh Malaysia Plan.

Malaysia's 2016 Budget has shown that government emphasized on developing more affordable houses for public as one of the highest concerned aspects in the budget. 1.6 billion ringgit is allocated for 175,000 affordable housing projects. Furthermore, in order to tackle the housing affordability issue, government introduced a First Home Scheme that offers up to 100% financing or zero down-payment to first home buyers below 35 years old with an allocation of RM200 million. Their aims are to help affordable home buyers with their down payments on their first house purchase. Under Perumahan Rakyat 1Malaysia (PR1MA), Rumah Mesra Rakyat, Perumahan Penjawat Awam 1Malaysia (PPA1M) and Program Perumahan Rakyat, a total of 317,000 units affordable homes have been planned.

Despite all the initiatives taken by the government, there exist an argument regarding the housing price and housing affordability in Malaysia. For example, only half of the 1,642 applicants were found to be eligible to receive a loan from the bank to buy a house under the First Home Scheme [2]. In terms of housing construction, only 1,382,917 units were construction in the 6th and 7th Malaysia Plan while approximately more that 8million units is needed in order to achieve vision 2020 [3]. As the government is tackling housing affordability issues, it is important to analyse the effectiveness of the approach from consumer or end-user's point of view. Public opinion and reviews towards the initiatives either appreciated or criticized can be an excellent way for the government to plan for an improvement in the future.

It is becoming a trend now that people love to voice out what they think and feel via social network. With the rapid growth of social media such as Twitter, blogs and Facebook, the number of customer comments available to government, organizations and decision makers interested in obtaining public insights becomes larger than ever. It may no longer be necessary to conduct surveys, questionnaires, organize focus groups or employ external consultants in order to find consumer opinions about certain matter because social network platform can provide such information. According to statstool.com, Facebook has a Google PageRank of 9/10 with their daily unique visitors approximately over 36 million while Twitter also has been reported to have 36.1 million visitors daily. This kind of scenario gives a very good opportunity to an organizations or government to get feedback and to understand people's feeling and opinion over any issues. Sentiment analysis, also known as opinion mining, grows out of this need.

We are proposing a study that focuses on opinion expressions that convey people's positive or negative sentiments towards government initiatives in dealing with housing issues. Sentiment Analysis (SA) is a discipline that has seen a lot of research has been conducted since about 2000 [4]. The main reason for this is the proliferation of social media and its tools (e.g. Twitter, Facebook, LinkedIn, etc.) that has made the accessibility to information about people's feeling and thought about things more readily available to the masses.

There are two major goals of this paper. First, we want to do a proper analysis on each aspect of housing projects in order to get the real public opinion in property situation in Malaysia and second, to assist the organizations/government in decision making process by visualizing the results of the analysis in real-time basis. This work can be a good source for the decision maker to understand what people think about their efforts and could make a big difference in making wise decision in the future.

The rest of the paper is organized in sections. Section 2 presents the Problem Description. In Section 3 we cover the related work in Sentiment Analysis and in Section 4 we present the conclusion and discussion. Finally we present the future work in Section 5.

II. PROBLEM DESCRIPTION

Previous sentiment analysis application has been extensively studies in various domains such as mobile and movie reviews, product reviews, forecasting market movements and election results. To our knowledge, none of previous study has been done in housing domain. As housing is the essential and important issues in daily life, it does motivate us to apply sentiment analysis in this area.

Sentiment analysis can be conducted in three levels: document, sentence or feature/aspect based. Our proposed study is to analyse the sentiment at feature or aspects level. Feature-level sentiment analysis yields very fine-grained sentiment information which can be useful for applications in property development industry. Since our study is focusing on housing, feature based is suitable as we're not only analysing the project as a whole, but each property features such as price, location and design will be analysed as well. This defined why feature based is the best in this case.

Majority of previous studies on sentiment analysis have focused on analysing English language, for which a large amount of resources and tools for text mining and natural language processing are available. In addition, sentiment analysis in the Malaysian environment had not yet been thoroughly targeted by the research community. We have seen the need for more studies on sentiment analysis in other languages, such as Chinese, Spanish and Malay, are growing. As this study will focus on analysing opinion from Malaysian and both Malay and English are prominent for most Malaysian, this motivates us to conduct research for both languages.

The purpose of this study is not only understanding people emotions and thoughts, but also to assist decision maker in planning for improvement in the future. The idea of visualizing the result grows out of this need. Figure 1 shows the overall process of our proposed study.

III. RELATED WORK

This section illustrates other similar work related to analyzing sentiments.

A. Sentiment Analysis Application in Social Network

In recent years, an extremely large growth of the Internet usage has been observed. Internet has changed the way we communicate, search for information and do our business. It even affects our daily life. A number of new Web sites, technologies, and tools have been developed for use on the Internet.
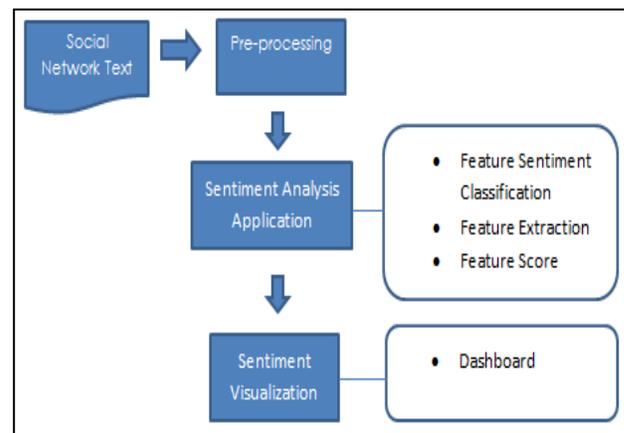

Figure 1: Overall process of proposed study

Among them, many social networking such as Facebook and Twitter have been developed and became very popular. Facebook is a social networking web site which was founded in 2004 and since then it has grown into a global network with 1.23 billion monthly active users [5]. On the other hand, Twitter founded in 2006, which rapidly gained global popularity with 271 million monthly active users [6].

The application of sentiment analysis over big data has leads to a lot of business and social benefits especially in decision making process [7]. Previous sentiment work in Twitter and Facebook data include the measurement and prediction of various domains like the prediction of election results [8], monitoring students thought in education [9] and many more.

B. Sentiment Analysis Levels

Sentiment Analysis is classified in three levels; document, sentence and feature/aspect level.

- Document level - Among the research in document level classification has been done by Pang [10] and Mullen [11]. Pang has used supervised machine learning method to classify movie reviews while Mullen had combined semantic orientation words defined by Turney [12] and few information he gained from Wen and thesaurus. They both have achieved 85% accuracy with the sentiment classification.

- Sentence-level - It classifies sentiment expressed in each sentence. If the sentence is subjective, it classifies it in positive or negative opinions. The research done by Kudo [13] has used the sentence level classification approach. It considered word dependency trees as features for sentence-wise sentiment polarity classification. On the contrary, the study conducted by Turney [12] determined the relationship between a polarity-unknown word and a set of selected manually seeds for classifying the polarity-unknown word into positive or negative class. A framework introduced by Jusoh and Hejab [14], has used possibility theory and fuzzy sets for the semantic extractions at sentence level. Another study done by Hu et al.[15] have extract

sentiment polarities by using expressions such as "fast but inaccurate" or "beautiful and smart". Those researches focused on the entire sentence to perform the analysis, they did not focus on specific features of the review.

- Feature based or aspect-level - Feature based classifies the sentiment with respect to the specific feature of entities. Users can give different opinions for different aspects/features of the same entity.

Previous works on feature based sentiment analysis by Liu [4] attempted to tackle sentiment and semantic labeling using different approaches such as sequence labeling [16], syntactic patterns [17], topic models [18]. While some works first separate the semantic and sentiment information and then label them [19], some other previous works presented joint models for joint semantic and sentiment labelling [20]. Vector representations for words and phrases have been found useful for many NLP tasks [21, 22] Wang et al. [23] have done a feature level sentiment analysis approach for product reviews in Chinese language. They have considered both modification of relationships between words and punctuations in review texts.

According to Liu [4], both the document level and sentence level analyses do not discover what exactly people liked or not. Studying the opinion text, mainly feature level, is extremely challenging. For the ordinary user, it is too complex to analyze opinions about object and object features in the great number of social networks available on the Web.

C. Language/Corpora

There are few related research has been done for sentiment classification models in Malay. Samsudin et al. [24] introduced a model that uses a method, where noisy texts that are available in Mixed Malay Language texts are normalized with the creation of the Malay Mixed Text Normalization Approach. Puteh [25] concentrated on preprocessing process for stemming Malay text, with the use of the Reverse Porter Algorithm and Backward Forward Algorithm. Furthermore, Puteh [25] also uses the artificial immune network to retrieve opinions from Malay newspaper articles. To our knowledge, none of previous research has been conducted on the combination of Malay and English language. In addition, most of the sentiment classification was done on structured data like newspaper, while our research will focus on social media platform which are informal and unstructured.

D. Sentiment Visualization

Visualizing the results of the analysis can provide a clear picture and useful insights about topic of interest to the public. Das et al. [26] propose a 5Ws method (who, what, where, when and why) of sentiment visualization in order to provide a complete contextual information in regard to sentiment analysis. In addition, Das et al. [26] state that sentiment aggregation is important in order to minimize mental load and to give better understanding. Dasgupta and Ng [27] contrast two aggregation levels: topic-wise vs sentiment-wise. Fukuhara et al. [28] propose temporal sentiment analysis that aggregates sentiment in both kinds of aggregations.

IV. CONCLUSION AND DISCUSSION

Since Tenth Malaysia Plan, housing issues has become the main concern by the government. In order to provide more affordable housing for the public, various initiatives has been done and still continued until now. It is important for the government to know people reactions towards those initiatives either being appreciated or criticized. This paper proposed the application of sentiment analysis to understand people's feeling and thoughts. Sentiment analysis is a field of study that analyses people's sentiments, attitudes, or emotions towards certain entities. Social network streams are selected to be our sources of sentiment. Our work proposes a framework for the application of sentiment analysis in housing projects. Three levels of sentiment analysis have been presented which are document, sentence and feature level. This paper tackles a problem of sentiment classification at feature level in Malay and English language. The aim of this research is to provide an effective analysis in capturing the real voices of consumers in regards to property development industry. In addition, the result obtained from the analysis will then be visualized in order to ease the government or any organizations in the decision making process

V. FUTURE WORK

Our future work aims towards the implementation of this proposed framework and to do the validation using real world datasets which will be extracted from various social media platforms such as Facebook and Twitter. Our focus will be on developing both Malay and English sentiment corpus, investigating the implementation of feature based sentiment analysis and to visualize the results of the analysis in the best way possible.

REFERENCES

[1] Abdullah, M.R.B. Real Estate Market Structure: Industrial Organization Perspectives. IJMS. 22(1): 57-72. 2015

[2] Osmadi A, E. M. Kamal, H. Hassan, and H. A. Fattah. Exploring the Elements of Housing Price in Malaysia. Asian Social Science. 11: 24-26. 2011.

[3] Nawi, M.N.M., Nifa, F.A.A., Abdullah, S. and Yasin, F.M., 2007, November. A preliminary survey of the application of Industrialised Building System (IBS) in Kedah and Perlis Malaysian construction industry. In Conference on Sustainable Building South East Asia (Vol. 5, No. 7). 2007.

[4] Liu, B. Sentiment analysis and subjectivity. Handbook of Natural Language Processing, Second Edition. CRC Press, Taylor and Francis Group. 2012.

[5] "Facebook: 10 Years Of Social Networking, In Numbers". the Guardian. N.p. 2014. Web. 9 Feb. 2016.

[6] Craig, S. "170+ Amazing Twitter Statistics". DMR. N.p. 2014. Web. 9 Feb. 2016

[7] Borth, D., R. Ji, T. Chen, T. Breuel, and S.-F. Chang. Large-scale visual sentiment ontology and detectors using adjective noun pairs. ACM international conference on Multimedia. 223–232. 2013.

[8] Hodson, H. Twitter hashtags predict rising tension in Egypt. New Scientist. 219(2931): 22. 2013.

[9] Ortigosa, A., J. M. Mart´ın, and R. M. Carro. 2014. Sentiment analysis in facebook and its application to e-learning. Computers in Human Behavior. 31: 527–541. 2014.

[10] Pang, B., L. Lee, and S. Vaithyanathan. Thumbs up? Sentiment Classification using Machine Learning Techniques. 7th EMNLP. 79-86. 2002.

[11] Mullen, T., N. Collier. Sentiment Analysis using Support Vector Machines with Diverse Information Sources. 9th EMNLP. 412-418. 2004.

[12] Turney, P. Thumbs up or thumbs down? semantic orientation applied to unsupervised classifcation of reviews. 40th ACL, pp. 417-424. 2002.

[13] Kudo, T., Y. Matsumoto. A Boosting Algorithm for Classification of Semi-Structured Text. 9th EMNLP, Vol. 4, pp. 301-308. 2004.

[14] Jusoh, S. and Al Fawareh, H.M., 2011. Semantic extraction from texts. InProceedings of International Conference on Computer Engineering and Applications IPCSIT, Singapore. 2011.

[15] Hu, X., J. Tang, H. Gao, and H. Liu. Unsupervised sentiment analysis with emotional signals. International Conference on World Wide Web. 2013.

[16] Yang, B., and C. Cardie. Joint inference for fine-grained opinion extraction. Annual Meeting of the Association for Computational Linguistics. 2013.

[17] Zhou, X., X. Wan, and J. Xiao. Collective opinion target extraction in Chinese microblogs. EMNLP, pp. 1840–1850. 2013.

[18] Lu, Y., Q. Mei, and C. Zhai. 2011. Investigating task performance of probabilistic topic models: an empirical study of plsa and lda. Information Retrieval, 14(2):178–203. 2011.

[19] Choi, Y., and C. Cardie. Hierarchical sequential learning for extracting opinions and their attributes. ACL 2010 Conference Short Papers, pp. 269–274, Stroudsburg, PA, USA. 2010.

[20] Jo, Y., and A. H. Oh. 2011. Aspect and sentiment unification model for online review analysis. WSDM, pages 815–824, New York, NY, USA. 2011.

[21] Bansal, M., K. Gimpel, and K. Livescu. 2014. Tailoring continuous word representations for dependency parsing. ACL, pp. 809–815, Baltimore, MD, USA. 2014.

[22] Chen, Y. N., and A. I. Rudnicky. Dynamically supporting unexplored domains in conversational interactions by enriching semantics with neural word embeddings. 2014 Spoken Language Technology Workshop, pp. 590– 595. 2014.

[23] Wang, H., L. Liu, and W. Song. Feature-Based Sentiment Analysis Approach For Product Reviews. Journal of Software. 9(2). 2014.

[24] Samsudin, N., M. Puteh, A. Hamdan, M. Zakree, and A. Nazri. Immune Based Feature Selection for Opinion Mining. World Congress on Engineering. 120-127. 2013.

[25] Puteh, M., N. Isa, S. Puteh, and N. A. Redzuan . Sentiment Mining of Malay Newspaper (SAMNews) Using Artificial Immune System. Congress on Engineering. 2013.

[26] Das, A., S. Bandyopadhyay, and B. Gamback.. Sentiment analysis: What is the end user's requirement? WIMS 2012. 35:1–35:10.2012

[27] Dasgupta, S., and V. Ng. 2009. Topic-wise, sentiment-wise, or otherwise?: Identifying the hidden dimension for unsupervised text classification. EMNLP 2009. 580–589. 2009.

[28] Fukuhara, T., H. Nakagawa, and T. Nishida. Understanding sentiment of people from news articles: Temporal sentiment analysis of social events. ICWSM 2007.