Automatic Sentiment Analysis from Opinion of Thais Speech Audio

Preedawon Kadmateekarun
Faculty of Science and Technology
Suan Sunandha Rajabhat University
Bangkok, Thailand
preedawan.ka@ssru.ac.th

Sumitra Nuanmeesri
Faculty of Science and Technology
Suan Sunandha Rajabhat University
Bangkok, Thailand
sumitra.nu@ssru.ac.th

***Abstract*— Automatic classification of sentiment is widely used in academia and industry by several techniques. This paper aims to develop a method of sentiment analysis for Thais customers to identify the different notions into two opinions (positive or negative) to consume the products. These opinions are represented by text that is derived from the Thais speech audio content in social media especially video reviews about beauty product. Then, this work implements the model by the Naïve Bayes text classification. The results could be demonstrated that the method can provide more effectiveness and satisfactory accuracy for automatic sentiment analysis.**

***Keywords*— sentiment analysis, Thais speech audio, Naïve Bayes, text classification**

I. Introduction

Currently, analyzing people's opinions by using sentiment analysis towards services and products is getting more attentions. It is a part of Natural Language Processing (NLP) that its procedure focuses on analyzing and examining opinions of people with regard to messages as text that they posted on the internet. The opinions showed their feeling on things. It can be determined whether the opinions regarding the features are positive (Good) or negative (Bad) [1] [2] [3] [4] [5]. Nowadays, people's opinions analyzing techniques are being more interesting, as well as being applied widely including video reviews related to beauty. These videos were used as tools in analyzing customer opinions that they like beauty products both efficiency and price while they use or not. If marketing operators are interested in these opinions, they can use as feedbacks to improve their product and enhance their customers' satisfaction because this kind of messages can affect to other customer too. This research presents the opinions of reviewers or customers who use beauty products and post some video reviews on YouTube especially in Thais speech audio. These opinions were analyzed by using key words "like" or "do not like".

II. Related Works

Most researches related to sentiment analysis were applied to customer opinions which it was called "Product Review" [1] [2] [3] [4] [5] toward the services or products of organizations or other companies. Normally, only the rating data from questionnaires could not identify problems or customers' opinion toward services or products really; for example, after customers bought a digital watch, they might score 4 out of 5 for that watch; however, the questionnaires related to products might have not covered what customers really needed or wanted from the product. So, customers might have written some opinions toward that products, and posted on other places such as Blog, Twitter, or Facebook [6] [7]. In fact, those opinions can be applied to improve the services or products by product owners or service providers [5].

Recently, technology helping in sentiment analysis is becoming influence on many organizations in a field of product distributors, education providers, and medical services. This kind of technology has some costs and were included within the commercial website or some marketing software such as Customer Relationship Management (CRM) in many companies or organizations which could ease of use to analyze people's opinions and led to solve problems of products or services rapidly [5].

A. Text-Based Sentiment Analysis

This technique was developed especially for sentiment analysis concentrated on text processing method. It consists of either rule-based classifiers that use opinion-orientation, or data-driven methods that forecast the availability of a large dataset annotated for polarity. The General Inquirer (GI) was one of the first lexicons used in polarity analysis; since its introduction, many methods have been developed to automatically identify opinion words [2] [3], *n*-grams, and more linguistically complex phrases [4] [5]. For data-driven methods, one of the most widely used datasets is the MPQA corpus [6], which is a collection of news articles manually annotated for opinions. Other datasets are also available, including two polarity datasets covering the domain of movie reviews [7] [8], and a collection of newspaper headlines annotated for polarity [9]. More recently, multi-domain [10] and multilingual [11] resources have also become available.

Besides, these techniques and other relevant resources, there is another work that concerned with the automatic identification of subjectivity and sentiment in text. It mostly focuses on texts such as online reviews [2] [7], news articles

978-1-4673-7670-9/15/$31.00 ©2015 IEEE

[12], web blogs [13], or Twitter [14]. Tasks such as cross-domain[15] or cross-language [11] [15] portability have also been addressed. Despite, there is a little progress on the processing of sentiment in text; it has been done in term of extending the applicability of sentiment analysis to other modalities, such as speech, gesture, or facial expressions. However, two exceptions must be considered. First, in the research reports elsewhere [16], speech and text are analyzed together for the purpose of subjectivity identification. Second, it did not address other modalities such as visual cues, and did not address the problem of sentiment analysis.

B. *Audio Sentiment Analysis*

More research addressing the multimodal fusion of language, acoustic features, and visual gestures, such as the Video Information Retrieval Using Subtitles (VIRUS) project that uses all three modalities to perform video retrieval and research on audio and/or visual emotion analysis. Some recent surveys discuss dimensional and categorical affect recognition such as Martin Wollmer defines a novel algorithm based on a combination of audio-visual features for emotion recognition. Mihalis Nicolaou proposes the use of Output- Associative Relevance Vector Machine (OA-RVM) for dimensional and continuous prediction of emotions based on automatically tracked facial feature points.

Moreover, researches considered individual audio or visual modalities, there's also a growing of work concerned with audio-visual emotion analysis. The features used by novel algorithms are usually low-level features, such as tracking points for collecting visual data, or audio features like pitch level. In addition to work that a challenge was organized focusing on the recognition of emotions using audio and visual cues, which included sub challenges on audio-only, video-only, and audio-video, and drew the participation of many teams from around the world. And also related research is the multimodal integration of opinion mining and facial expressions, which can be successfully used for the development of intelligent affective interfaces.

It's more important to note that multimodal emotion recognition is different from multimodal sentiment analysis. Although opinion polarity is often correlated to emotional valence (as used in the datasets for audio-video emotion analysis), these concepts are quite different. For example, someone can be smiling while at the same time expressing a negative opinion, which makes multimodal sentiment analysis a complex and challenging research direction.

Recently research, in a pre-study on 47 English videos [17] and 76 Spanish videos on YouTube, it has been shown that visual and audio features can complement textual features for sentiment analysis. In our work, we use a new dataset focusing on Thais speech audio which it was extracted from video reviews related to beauty products, and draw summary features at video level. Moreover, we show that multimodal sentiment analysis can be effectively used for sentiment analysis on different languages.

III. RESEARCH METHODOLOGY

This part explored the research methodology which it was including the details in each procedure of this research. Fig. 1 shows that automatic sentiment analysis was divided into four main parts, there are: sentiment modeling, audio extraction, speech to text and word segmentation, and automatic sentiment analysis engine.

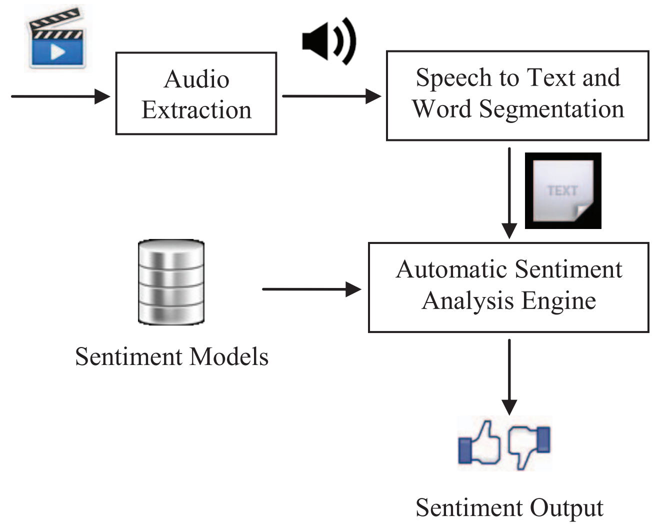

Figure 1. The methodology for automatic semtiment analysis.

A. *Sentiment Modeling*

This model was consisted of three steps as the following:

First step, document preprocessing, this began by tokenization which was to classify the sentences and separated to smallest unit in languages, it was called "word". In general, these words could be done by using the dictionary approach based on the Longest Matching Algorithm. For example, the sentence "I eat apple." could have been cut to be I/eat/apple. The next step was stop words which were the process in cutting any word or symbols that could frequently find in many messages. After cutting these words off, the sentence was still meaningful, although a word or a symbol did not affect to the classifier. In Thai language, the stop word could be a conjunction or a short function word. The tokenization was very important in classifying the automatic sentiment modeling as it could save costs in complying.

Second step, document representation was the relationship between "word" and "document" in form of vector by focus on "word" or "phrase" in order to learn and test with the artificial neural network. The statistics such as frequency of word or phrase and grammar rules are success keys to find relationship in document. In the representation, Vector Space Model (VSM) or Bag of Words (BOW) was used to reveal the relationship between the document and the words in form of 2 dimensions. Replaced the document by BOW was to determine "word" in the document by w_{ij}. Therefore, any document j could be replaced as in (1)

$$d_j = (w_{11}, w_{12}, w_{13}, \ldots, w_{ij}). \tag{1}$$

This research used *tf-idf* (Term Frequency – Inverted Document Frequency) in order to show the relationship between "word" and document. This *tf-idf* was the tool to create the representation of document in form of vector which was applied to classify document to the designated categories. *tf* was the frequency in the document and *idf* was the invert of the document or on the other word it was called Inverted Document Frequency. This could be found by apply this algorithm:

$$idf = 1 + log(\mathrm{N}/df). \tag{2}$$

N was the total number of documents in the group and df was the number of documents that "word" existed.

$$tf - idf = tf \times idf. \tag{3}$$

tf-idf was the way how to find the representation of vector which could be used to find out in the document. This was easy and efficient in classify the document.

Last step, Naïve Bayes analysis model was a algorithm that was widely used in classifying documents, and achieved the good result. To find out the number of Naïve Bayes, this could be done by beginning with Instance x in form of vector as the following: $<a_1, a_2, \ldots, a_n>$ The target value was that any values of Instance were V and V was the target value that meant the number of data. Naïve Bayes was applied in Text Classification and it worked very well and this was similar to other methods. Thus, the researcher chose this method in this research according to efficiency and simplicity compared with other methods.

The probability belonged to the data which was assigned to V_j. For the data that had quality n then $X = \{a_1, a_2, \ldots, a_n\}$ or $P(a_1, a_2, \ldots, a_n)$.

$$P(a_1, a_2, \ldots, a_n) = \prod_{i=1}^{n} P(ai \mid vj). \tag{4}$$

$\prod_{i=1}^{n}$ meant the multiply value of $P(a_i \mid v_j)$ when i *and* j valued 1, 2, 3, …, n.

The easy way to learn about Bayes could be as the following:

(1) Find the probability of words in each group by applying $P(a_1, a_2, \ldots, a_n \mid v_j)$ and multiply by the probability of that group $P(v_j)$ which would be V_{NB}
(2) Compare the value between the highest probability and assigned group as the following algorithm:

$$V_{NB} = argmax\ P(v_j) \times \prod_{i=1}^{n} P(ai \mid vj). \tag{5}$$

B. Audio Extraction

The audio extraction is the process to extract or rip the audio track from video clip or movie. In usually, this process will be separate the video clip into two parts: audio track and movie track or visualize information. It has much software to extract the audio track from video clip or YouTube such as YouTube-dl, user can extract audio with some command line.

C. Speech to Text and Word Segmentation

This part described to process for derive the audio track to sentences or messages as text in speech to text step. After that these texts to be cut out into words in word segmentation step. On the other word could be called "Transcription". The transcriber software was used in this part to perform only the audio track.

D. Automatic Sentiment Analysis Engine

The last part explored to sentiment analysis process the opinions from texts which it were separated into words by word segmentation process. These words were cutting the stop words off in the messages or sentences. After getting "word" from messages, words data were applied in analyzing to classifying the data according to algorithm (5). From algorithm (5), the probability was assessed for class by class between positive and negative. So, P(class) value is 0.5. For example:

*D*NEW : *ฉันคิดว่าลิปสีนูทสุดยอด (I think that's the ultimate nude lip.*)

V_{NB} = P(+) x P(สุดยอด|+) x P(ลิปสีนูท|+)
= (0.5) x (0.602) x (0.1505) = 0.0453005

Checked in *class* = "*Negative*"

V_{NB} = P(-) x P(สุดยอด|+) x P(ลิปสีนูท |+)
= (0.5) x (0) x (0.1505) = 0

From above result, it clearly revealed that *D*NEW had probability value at *class* = "*Positive*" more than *class* = "*Negative*". Therefore, it assumed that *D*NEW was classified in positive or customer like this sentence in the movie clip.

IV. Experimental Results

This research used Recall, Precision, and F-measure to assess the efficiency of computing and accuracy of the analysis processing. When running the test in classifying the 150 video samples by applying the sentiment analysis on different video in order to create a model, the value of Recall, Precision, and F-measure was 40, 50, and 60 respectively. Table I shows the results of automatic sentiment analysis with 150 video samples. The result of sentiment analysis was satisfaction because the average results had Recall value at 0.833, Precision value at 0.877 and F-measure value at 0.841.

TABLE I. Experimental Results

Model Types	Recall	Precision	F-measure
40 video samples per group	0.82	0.85	0.827
50 video samples per group	0.83	0.88	0.836
60 video samples per group	0.85	0.90	0.859
Average	0.833	0.877	0.841

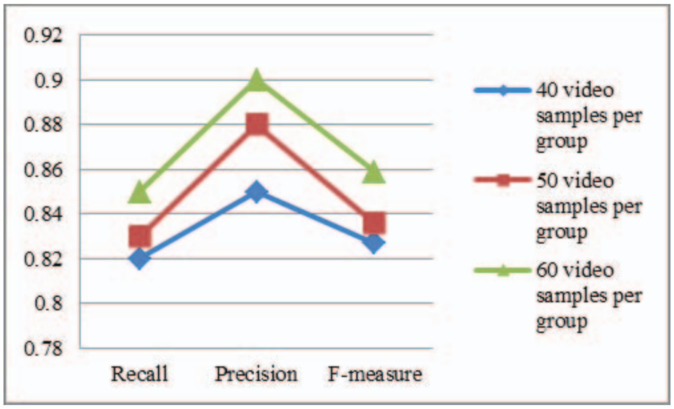

Figure 2. Comparison of model types.

Fig. 2 shows that the model has developed effective. The results of the comparison sample of video in each group are different; the higher is the better direction.

V. CONCLUSION

This research analyzed the automatically sentiment the opinions from videos were posted on YouTube by reviewers or customers who use beauty products especially in Thais speech audio. This was done on both positive and negative. This research applied Naïve Bayes which was used as the main algorithm in creating a model for sentiment analysis in the processes of measuring the efficiency of classifying the opinions of consumers. This research used Precision, Recall, and F-measure in testing the efficiency of the computing models, and found that the average value of Recall was 0.833, Precision was 0.877 and F-measure was at 0.841. Even though the result was satisfaction, it can be concluded that the model was developed with effectively.

REFERENCES

[1] M.S. Neethu and R. Rajasree, "Sentiment Analysis in Twitter using Machine Learning Techniques," Computing, Communications and Networking Technologies (ICCCNT), 2013 Fourth International Conference, July 2013.

[2] J.S. Modha, H. G. S. Pandi, S.J. Modha, "Automatic Sentiment Analysis for Unstructured Data", International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, 2013.

[3] S. Modak and A.C. Mondal, "A Study on Sentiment Analysis," International Journal of Advanced Research in Computer Science &Technology, vol. 2, 2014.

[4] H. Gomes, M. de Castro Neto, R. Henriques, "Text Mining: Sentiment analysis on news classification," The 8th Iberian Conference on Information Systems and Technologies (CISTI), June 2013.

[5] J. Polpinij and A.K. Ghose, "An Ontology-Based Sentiment Classification Methodology for Online Consumer Reviews," Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, pp. 518-524, 2008.

[6] A.C. Lima and L.N. de Castro, "Automatic sentiment analysis of Twitter messages", Fourth International Conference on Computational Aspects of Social Networks, pp. 52-57, 2012.

[7] L. Jiang, M. Yu, M. Zhou, X. Liu and T. Zhao, "Targetdependent Twitter sentiment classification," Proceeding HLT' 11 Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 2011.

[8] K. Balog, G. Mishne and M. de Rijke, "Why Are They Excited? Identifying and Explaining Spikes in Blog Mood Levels," Proc. 11th Meeting of the European Chapter of the Assoc. Computational Linguistics, ACL, 2006, pp. 207–210.

[9] P. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews," Proc. 40th Ann. Meeting of the Assoc. Computational Linguistics, ACL, 2002, pp. 417–424.

[10] M. Taboada et al., "Lexicon-Based Methods for Sentiment Analysis," J. Computational Linguistics, vol. 37, no. 3, 2011, pp. 267–307.

[11] E. Riloff and J. Wiebe, "Learning Extraction Patterns for Subjective Expressions," Proc. Conf. Empirical Methods in Natural Language Processing, 2003, pp. 105–112.

[12] H. Takamura, T. Inui, and M. Okumura, "Latent Variable Models for Semantic Orientations of Phrases," Proc. European Chapter of the Association for Computational Linguistics, ACL, 2006, pp. 201–208.

[13] J. Wiebe, T. Wilson, and C. Cardie, "Annotating Expressions of Opinions and Emotions in Language," Language Resources and Evaluation, vol. 39, nos. 2–3, 2005, pp. 165–210.

[14] B. Pang and L. Lee, "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts," Proc. 42nd Meeting of the Assoc. Computational Linguistics, ACL, 2004, pp. 271–278.

[15] A. Maas et al., "Learning Word Vectors for Sentiment Analysis," Proc. Assoc. Computational Linguistics, ACL, 2011, pp. 142–150.

[16] C. Strapparava and R. Mihalcea, "Semeval-2007 Task 14: Affective Text," Proc. 4th Int'l Workshop on the Semantic Evaluations, ACL, 2007, pp. 70–74.

[17] J. Blitzer, M. Dredze, and F. Pereira, "Biographies, Bollywood, Boom-Boxes, and Blenders: Domain Adaptation for Sentiment Classification," Proc. Assoc. Computational Linguistics, ACL, 2007, pp. 187–205.

[18] C. Banea, R. Mihalcea, and J. Wiebe, "Multilingual Subjectivity: Are More Languages Better?," Proc. 23rd Int'l Conf. Computational Linguistics, ACL, 2010, pp. 28–36.

[19] N. Godbole, M.Srinivasaiah, and S. Sekine, "Large-Scale Sentiment Analysis for News and Blogs," Proc. Int'l Conf. Weblogs and Social Media, 2007; http://icwsm.org/papers/3--Godbole-Srinivasaiah-Skiena.pdf.

[20] L. Jiang et al., "Target-Dependent Twitter Sentiment Classification," Proc. Assoc. Computational Linguistics, ACL, 2011, pp. 151–160.

[21] X. Wan, "Co-Training for Cross-Lingual Sentiment Classification," Proc. Joint Conf. Assoc. Computational Linguistics and the Int'l Joint Conf. Natural Language Processing, ACL, 2009, pp. 235–243.

[22] S. Raaijmakers, K. Truong, and T. Wilson, "Multimodal Subjectivity Analysis of Multiparty Conversation," Proc. Conf. Empirical Methods in Natural Language Processing, ACL, 2008, pp. 466–474.

[23] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[24] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[25] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.