Using Good and Bad Sentiment to Differentiate Good and Bad News

Sasiphan Nitayaprapha

Faculty of Liberal Arts, Thammasat University

DOI: 10.29322/IJSRP.14.05.2024.p14922
10.29322/IJSRP.14.05.2023.p14922

Paper Received Date: 14th April 2024
Paper Acceptance Date: 12th May 2024
Paper Publication Date: 22nd May 2024

Abstract- News consumption is critical to human's mental health In addition, bad news has more impact than good news. In general, humans consume more bad news than good news. Bad news is usually distributed faster than good news. Consumption of bad news could have long-lasting negative effects on human's well being, Bad news could lead to anxiety, stress, and depression.
The research aims to provide a filtering guideline to differentiate good news and bad news. The research employs AI for Thai, an NLP tool, to do word segmentation and sentiment analysis. The filtering system distinguishes good and bad news from good and bad sentiment. The data are collected from covid-19 news posts of the three main Thai news Facebook pages; Thairath – 17 millions followers, เรื่องเล่าเช้านี้ – 16 millions followers, and ข่าวเวิร์คพอยท์ 23 – 2.3 millions followers during 2020-2022. Based on covid-19 topics, the results shown that one media Facebook channel tends to be more positive, while another two is more negative, and neutral

Index Terms- sentiment analysis, NLP, text mining, data mining, news sentiment

I. Introduction

Social media have been a major source of information. A lot of social media contents report good and bad news related to recent events such as criminals, disasters, discoveries, elections, health, economies.

Research has shown that news played an important role in human's mental health [1]. In addition, bad news has more impact than good news. Moreover, humans tend to consume more bad news than good news. The recognition of bad news is increased because bad news is usually distributed faster than good [2]. Consumption of bad news could have long-lasting negative effects on human's well being, i.e. bad news induce anxiety and stress which could lead to depression [3]
These characteristics of bad news should be taken into consideration. Distinguishing good and bad news could help news readers to filter news and thus be to control the amount of bad news they consume. The aim of this paper is to provide a filtering system and guidelines which may help readers to filter out bad news. The research focuses on three main Thai news Facebook pages; Thairath – 17 millions followers, เรื่องเล่าเช้านี้ – 16 millions followers, and ข่าวเวิร์คพอยท์ 23 – 2.3 millions followers. The filtering system distinguishes good and bad news based on good and bad sentiment of the news respectively.

II. Literature Review

2.1 Natural Language Processing

Natural Language Processing (NLP) has become a major research area in artificial intelligence (AI). Thai NLP research has begun three decades ago [4]. The examples of Thai NLP language tools are TLTK [5] and AI for Thai. Thai language has unique features. Thai writing system lacks explicit delimiter of words. Additionally, Thai grammar allows all sentence structures whether subject-verb-object, subject-object-verb, and object-subject-verb. This uniqueness requires special NLP techniques.
In this paper, the researcher adopts AI for Thai for word segmentation and sentiment analysis of news posts. Word segmentation is essential task before further analyzing text.

According to Haruechaiyasak and Kongthon [6], the unknown words extracted from post could be classified in to four categories; insertion, onomatopoeia, transformation, and transliteration. The example of unknown insertion word is thankXXX. The example of unknown transformation is see u (see you). The word segmentation tools called LextoPlus is invented to cope with the insertion problems [6]. LextoPlus uses conditional random fields (CRFs) ,a statistical model, and dictionary to eliminate repeated character tokens. An inaccuracy elimination might occur if a word is not listed in dictionary.

TLex+ is another word segmentation tool invented by Kongyoung et al. [7]. TLex+ uses CRP and machine learning to identify characters whether it is beginning or intra word character. Tlex+ employs a list of terms and needs to be trained with characters and

types of characters. It is claimed that the tool outperforms LextoPlus.

2.2 Sentiment Analysis

Sentiment analysis is the analysis of texts to understand underlying emotion. It is a method that uses the natural language processing (NLP), text analysis and computational techniques to automate the extraction and classification of sentiment [8]. Sentiment analysis categorizes text emotions into polarity categories: positive, neutral, negative, or more types [9]. The examples of sentiment analysis research conducted on social media include those on Facebook, Twitter, News, and Products and services [10].

Many of previous studies develop sentiment dictionaries for specific domains. In addition, each word in the dictionary has its polarity values, positive, neutral, negative, or others. In fact, there are two approaches to score the sentiment, coarse(or binary) and fine-grained [11]. The coarse sentiment differentiates sentiment into positive, neutral, or negative polar [12] The fine-grained sentiment analysis classifies sentiment into many subgroups such as very negative, negative, neutral, positive, and very positive [13].

The sentiment value of texts is resulted from summation of the frequency of the composed words in the sentiment dictionary. Since, the sentiment analysis is based on a deployed sentiment dictionary, the available and reliable sentiment dictionary for a targeted textual domain is critical.

III RESEARCH METHODOLOGY AND DATA COLLECTION

The study uses bot to extract relevant data, and then carries out sentiment analysis by using Thai NLP tools call AI for Thai. Facebook-scraper bot is invented using Selenium and Python. Then AI for Thai is employed as a NLP tool to analyze Thai text. The system modules are as follows. The system architecture is depicted in figure 1.

A. Data Acquisition: this module extracts posts, along with date/ time of the posts, and exports posts to Excel file. News-scraping bot was created using selenium and python. The bot carries out four main tasks:

1) Open news Facebook fanpage.
2) Login with facebook account.
3) Extract relevant data.
4) Store data in Microsoft Excel file.

B. Data Preprocessing: data preprocessing deploys TLex+ to do word segmentation. The extracted new posts are shown by figure 2. The data comprised of two columns, datetime and news posts. The posts are collected in excel file format and cleansed to remove duplicates; for examples same posts on different dates, and posts which are not related to the covid-19. The news posts in 2020 are scarce and no longer exist. After data cleansing, the number of covid-19 news posts left for sentiment analysis is around 50 for each of three main Thai news Facebook pages.

C. Sentiment Categorization: this module uses s-sense to do sentiment analysis and categorize the post into positive neutral or negative. S-Sense is an application in AI for Thai developed for sentiment analysis on social media posts. The sentiment analysis is depicted in figure3.

News posts are collected from three main Thai news Facebook pages from 1 January 2020-31 December 2022. To maintain anonymity, the new channels are named news1, news2, and news3.

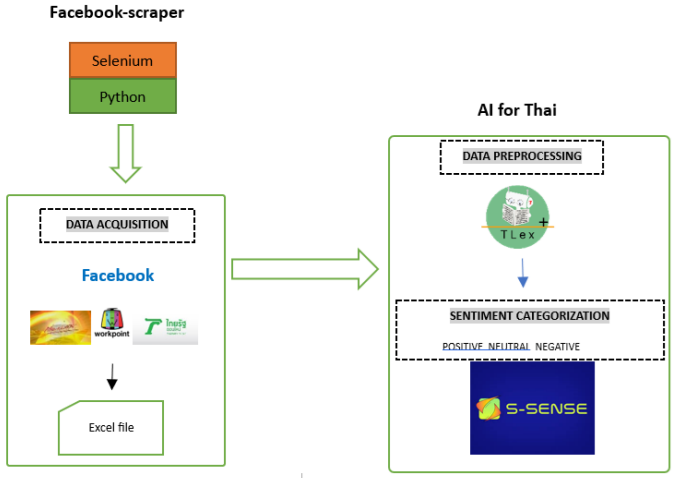

Figure1: System Architecture

Figure2: News posts extracted by the news-scraping bot

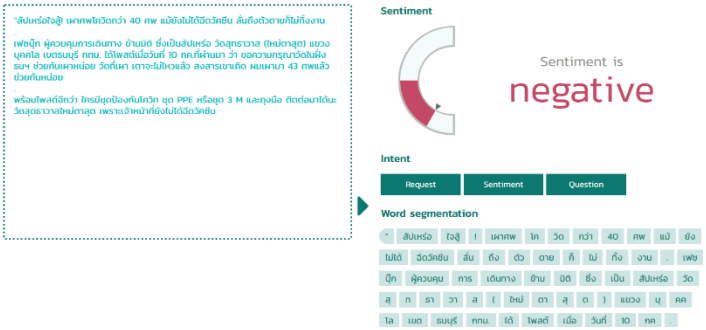

Figure3: Sentiment analysis by S-Sense

IV. RESEARCH RESULTS

The results of sentiment analysis of covid-19 news posts are presented in figure4. News2 gets the highest score in positive sentiment analysis (30.95%) and lowest score in negative sentiment analysis (40.48%), while news1 gets the highest score in negative sentiment analysis (53.06%) and news3 gets the highest score in neutral sentiment analysis (30.23%).

	Positive	Neutral	Negative	FP	FN
news1	30.61%	16.33%	53.06%	8.16%	0.00%
news2	30.95%	28.57%	40.48%	7.14%	9.52%
news3	23.26%	30.23%	46.51%	6.98%	2.33%

Figure4: Sentiment Analysis

To examine performance by taking FP and FN into consideration, the results are the same, i.e. news2 gets the highest score in positive sentiment analysis, news1 gets the highest score in negative sentiment analysis, and news3 gets the highest score in neutral sentiment analysis.

V. CONCLUSION

This paper investigates the good and bad sentiment of covid-19 news reported by the three main Thai news Facebook pages. The data collected are those that are exist and publicly available. Based on covid-19 topics, the results have shown that one media Facebook channel tends to be more positive, while the other two are more negative, and more neutral. In future studies, the research will expand investigation by including other topics. This allows to establish whether the news media channels are positive or negative when compare to the others, or provide consistent good and bad news, or are becoming more positive or negative over time. The findings could help providing guideline to filter sources of good news and bad news of news articles.

REFERENCES

[1] JK Kellerman, JL Hamilton, EA Selby, EM. Kleiman, "The Mental Health Impact of Daily News Exposure During the COVID-19 Pandemic: Ecological Momentary Assessment Study,". JMIR Ment Health, 2022, 25, 9(5), doi: 10.2196/36966. PMID: 35377320; PMCID: PMC9135112

[2] A. Fang, Z. Ben-Miled, "Does bad news spread faster?," International Conference on Computing, Networking and Communications (ICNC), 2017, pp.793-797.

[3] J. Baum, RA. Rahman, "Negative news dominates fast and slow brain responses and social judgments even after source credibility evaluation," NeuroImage, 2021, 244.

[4] V. Sornlertlamvanich, N. Takahashi, H. Isahara, "Building a thai part-of-speech tagged corpus (orchid)," Journal of the Acoustical Society of Japan (E), 1999, 20, 3, pp.189–198.

[5] W. Aroonmanakun, "Collocation and thai word segmentation," Proceedings of the Fifth Symposium on Natural Language Processing The FifthOriental COCOSDA Workshop, 2002, pp. 68–75

[6] C. Haruechaiyasak, A. Kongthon, "LexToPlus: A Thai lexeme tokenization and normalization tool," Proceedings of the 4th Workshopon South and Southeast Asian Natural Language Processing, 2013, pp.9–16.

[7] S. Kongyoung, A. Rugchatjaroen, K. Kosawat, "Tlex+: a hybrid method using conditional random fields and dictionaries for thai word segmentation," International Conference on Knowledge,Information, and Creativity Support Systems, 2015, pp. 112–125.

[8] A. Basant, M. Namita, B. Pooja, and G. Sonal, Sentiment Analysis Using Common-Sense and Context Information, Hindawi Publishing Corporation Computational Intelligence and Neuroscience, 2015.

[9]. Hussein Doaa Mohey El-Din Mohamed, "A survey on sentiment analysis challenges," Journal of King Saud University - Engineering Sciences, 2018, 30, 4, pp.330-338,

[10] G. Beigi, X. Hu, R. Maciejewski, H. Liu, "An Overview of Sentiment Analysis in Social Media and Its Applications in Disaster Relief," ,in: W. Pedrycz, SM. Chen (eds) Sentiment Analysis and Ontology Engineering. Studies in Computational Intelligence, 2016, 639, https://doi.org/10.1007/978-3-319-30319-2_13

[11] R. Kohtes, "From valence to emotions: How coarse versus fine-grained online sentiment can predict real-world outcomes," Anchor Academic Publishing, 2015.

[12] S. Baccianella, A. Esuli, F. Sebastiani, "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining," in Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), 2010.

[13] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, et al., "Recursive deep models for semantic compositionality over a sentiment treebank," Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2013, pp.1631–1642.

AUTHOR

Asst.Prof.Dr.Sasiphan Nitayaprapha
Ph.D. in Informatics (MBS,The University of Manchester)
Faculty of Liberal Arts, Thammasat University,
contact: sasiphan@tu.ac.th, annona.reticulata@gmail.com