Sentiment Analysis of Local Tourism in Thailand from YouTube Comments Using BiLSTM

Sanya Khruahong[1](✉), Olarik Surinta[2], and Sinh Cong Lam[3]

[1] Department of Computer Science and Information Technology, Faculty of Science, Naresuan University, Phitsanulok, Thailand
sanyak@nu.ac.th

[2] Multi-agent Intelligent Simulation Laboratory (MISL), Faculty of Informatics, Mahasarakham University, Mahasarakham, Thailand
olarik.s@msu.ac.th

[3] Faculty of Electronics and Telecommunication, VNU - University of Engineering and Technology, Hanoi, Vietnam
congls@vnu.edu.vn

Abstract. Currently, social networks, where people can express their opinion through content and comments, are fast developing and affect various areas of daily life; Particularly, some research on YouTube travel channels found that almost tourists and audiences leave comments about their attitudes to that place. Thus, mining the emotional recognition of comments through artificial intelligence can bring knowledge about the tourists' general view. This article analyzes the relationship(s) between social media use and its effect on community-based tourism in Thailand using the Social Media Sensing framework (S-Sense) as sentiment analysis and the Bidirectional Long Short-Term Memory (BiLSTM) methods to analyze the text comments. This research collected 51,280 comments on 114 Youtube Videos, which are tourist attractions in various provinces in Thailand. The approach categorizes attractions based on sentiment analysis of 60% or more, including restaurants, markets, historical sites, temples, or natural attractions. The results show that 67.51% of the 19,391 clean-processed comments were satisfied with those attraction places. Therefore S-Sense and BiLSTM models can be sufficient to analyze the attitude of comments about attraction places with from 43 to remain 33 keywords of 1,603 comments. Furthermore, the offered sentiment analysis method has higher precision, recall, and F1 scores.

Keywords: Social media analytics · Sentiment analysis · Text classification · Social media sensing · Bidirectional long short-term memory

1 Introduction

Tourism is considered an essential economic factor in Thailand since it can contribute a significant part in gross domestic product. In the past, the Thai government had the policy to attract many foreign tourists, but during the past year, a coronavirus epidemic brought a big problem. Thus, the govermonet encourage Thai to have more domestic

O. Surinta and K. Kam Fung Yuen (Eds.): MIWAI 2022, LNAI 13651, pp. 169–177, 2022.
https://doi.org/10.1007/978-3-031-20992-5_15

travels tostimulate the domestic economy. In addition, the government has policies to support local or community-based tourism [1, 2]. Local tourism [3, 4], or Tourism in the community, focuses on local lifestyles which the residents invite travelers to attend their location. Thus, tourism must adhere to environmental sustainability principles, society, and community culture, which can be broadcast through social media.

A Social Networking Site (SNS) [5, 6] is a website that connects people over the Internet and establishes an auxiliary channels for people to communicate easier over large distances. It can create a personal space, for example, Facebook, YouTube, or Twitter. The number of users has increased dramatically in the last ten years. Users primarily use for different purposes such as to show their thoughts and feelings on that moment [7, 8], and may be their new stories, places, or experiences. All of these personal thoughts and feelings are posted on the social media in the forms of texts, images, and videos. Mining comments' emotional propensity in the big data domain is valuable for learning network public sentiment. The analysis results may provide suitable guidance for tourist development.

Social Media Analytics (SMA) [9, 10] is one of data analytics that organizes different sections of evidence or information into categories according to purpose and the assumptions set. Social media analytic is one of the most popular forms of social media that can timely provide knowledge about different situations. SMA process consists of three following steps 1) collecting data, 2) analyzing the obtained interpretations, and 3) displaying the results in an easy-to-understand format.

Deep learning is a method of automating learning by simulating the behavior of human neural networks by overlapping multiple layers of neural networks [11]. The deep learning method is key to learning from data using a general-purpose learning procedure. Thus, deep learning can be developed for various tasks.

Therefore, the researchers are interested in applying social media analytics techniques and sentiment analysis (S-Sense) [12] to analyze local tourism data. The purpose is to use comments from YouTube channels to analyze attitude patterns in community tourist attractions. Moreover, deep learning is applied with Bidirectional Long Short-Term Memory (BiLSTM) models [13] to find relevant factors to recommend entrepreneurs in online marketing with data analysis for local tourism. The results mentioned will be both the process and the resulting form. The researchers hope it can be applied to other community attractions and lead to tourism industry development.

This paper proposes a sentiment analysis of YouTube's comments based on BiLSTM. The paper is organized as follows. Section 2 describes the related works of this research. After that, Sect. 3 introduces the methodology of Social Media Analytics. Section 4 shows the discussion and results. Finally, Sect. 5 concludes the paper and describes the future direction.

2 Related Work

This section describes the data analytics methods investigated in this paper. These methods are the most popular in the literature.

2.1 Big Data Analytics

Data analytics is a method that analyzes and extracts the raw data to make the conclusion which is the information. Big Data Analytics (BDA) [14, 15] is increasingly growing as a trending method that generates extensive data and provides a new opportunity for decision-making. Moreover, it can produce new understandings to improve results in many domains. Big data analytics systems [16] can provide proactive action for strategic decision-making to predict future volumes. Predictive analytics uses previous information to forecast user behavior and trends. Machine learning is applied to analyze data and make predictions. Although BDA is an excellent technique that can be used to analyze data, it may also require big data technology, which is costly.

2.2 Deep Learning

Deep learning is significantly advancing in artificial intelligence problems, which can be divided into feedforward neural network and recurrent neural network [11, 17, 18]. The performance of deep learning systems can usually be dramatically improved by merely scaling them up. With a lot more data and a lot more computation, they generally perform a lot better. Some deep learning techniques are applied to the text to understand the content of a sentence or article. So, it needs to understand in natural language processing.

2.3 Social Media Analytics (SMA)

SMA [19] collects information from websites, social media, and blogs, after which the data is analyzed to be used in business decision-making. Many people may not realize that this process includes fundamental analysis like retweets or likes, highlighting consumer insights on social media. The main objective of using SMA is to analyze the user-generated content and spread speed from social media [20].

SMA is the processe of analyzing, measuring, and predicting of digital interaction, relationships, topics, ideas, or content. The primary SMA process has three-stage: 1) Capture, 2) Understand 3) Present, which is called the "CUP framework [21, 22]." This research presents social media commentary on automobiles to adapt to the automobile industry and is suitable for use with other businesses. Some framework [23] describes two data collection approaches (semantically driven and user-driven) and two analytic methods (temporal analysis and corpus analysis). However, it may take more tools to access information, which may take too long to implement this technique.

2.4 Sentiment Analysis

Sentiment analysis [10, 24–27], which is a part of artificial intelligence, is clarified as the task of finding the opinions of users, attitudes, and emotions about specific information. Sentiment analysis is the core method of many social media monitoring systems and trend-analysis applications. Sentiment analysis is a specific text analysis task for valence identification and subjectivity analysis. In comparison, text analysis and text mining aim to analyze textural data and extract machine-readable facts. Social Media Sensing (S-Sense) [12] is a sentiment analysis model consistent with natural language processing, text mining, and sentiment analysis techniques. It can help a business or organization to recognize the various activities related to their organization such as monitoring satisfaction and public attitudes towards their product brands or services. Thus, the business and organization can better understand their customers' needs.

Sentiment analysis can be proposed in three levels: document, sentence, and aspect-based [25]. Document sentiment analysis aims to resolve the overall opinion on a particular entity and express positive or negative sentiments, such as a product and service. Next, sentence-level aims to classify the sentiment expressed in each sentence. Finally, Aspect-level seeks to organize the views concerning the specific aspects of entities. However, data sets are vital for review analysis, especially the primary sources that can support the efficiency of sentiment analysis. These reviews are essential to business holders because they can make business decisions based on analyzing user feedback about their products.

Sentiment analysis was developed in three ways: 1)sentiment analysis based on sentiment dictionary, 2)machine learning, and 3)deep learning. However, using the dictionaries technique need to find sufficient coverage of sentimental words and lack of domain words which is the big problems. Thus, much research used machine learning and deep learning methods in sentiment analysis and achieved good results.

The deep BiLSTM is used as the representation of the comment texts [13, 26, 28]. BiLSTM is an extension version of the LSTM methods in which two LSTMs are applied to the input data. In the first step, the sequence input is fed to the first LSTM. In the second step, the reverse of the input sequence is transferred into the second LSTM. The BiLSTM method improves the learning of long-term dependencies and accuracy performance. Text classification is one of the essential factors for sentiment analysis. Some languages with constraints used text classification, such as Arabic [29]. Significantly, Thai has a complicated pattern; therefore, this article uses BiLSTM technique to improve Text classification in Thai Sentiment analysis.

3 Methodology

This section presents the research methodology with the following steps. The framework has the following principles: 1) Data collection, 2) Sentiment analysis, and 3) Summarizing results as shown in Fig. 1.

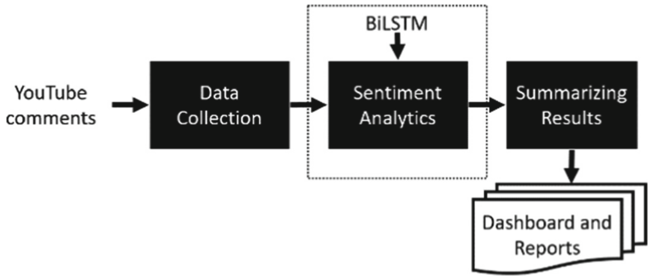

Fig. 1. Social media analytics framework

3.1 Data Collection

This research collected comment data from 114 clips on YouTube, which are tourist attractions in various provinces in Thailand. The 51,280 comments were collected with the service website for importing comments in Thai. We only collected those posts posted from April 17th, 2017, to May 15th, 2022. Text cleansing can be performed using simple Python code with "*AI for Thai*" API [30]. After being cleaned, the valuable comments reduces to 19,391 (Fig. 2).

Fig. 2. Original comment and text cleansing

3.2 Sentiment Analysis

Sentiment analysis further classified each text based on its sentiment, positive or negative, as shown in Fig. 3. S-Sense was applied with our comments data in the Thai language, which is analyzed as shown in Table 1. They display 13,091 positive comments and 6,300 negative comments. However, this paper only focuses on positive comments with a sentiment score of 60–100%.

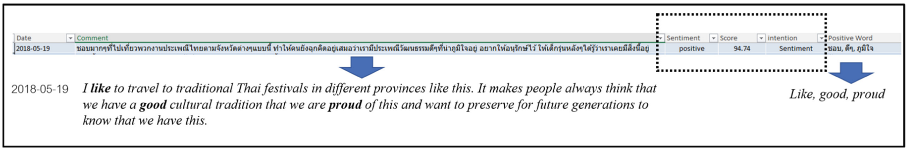

Fig. 3. Sentiment analysis results

Table 1. Number of sentiments from the collection by S-Sense

Sentiments	Comments	Score (Min)	Score (Max)
Positive	*13,091*	*60%*	*100%*
Negative	*6,300*	*−99.99%*	*−52.63%*
Total	***19,391***		

After S-Sense processing, 43 keywords categories were grouped from 2,631 comments, and we translated them from Thai into English shown in Table 2. The results show that some words may be exact and not present the attraction places, such as " ท่องเที่ยว" or " เที่ยว," they mean travel. Although this term is related to tourism, it does not reflect places or elements that can promote tourism. Those words should represent what tourists like, perhaps a place or something that can be considered a travel destination. Therefore, this paper needs to improve to get more suitable categories.

Table 2. Keywords of attraction with a positive sentiment

Keywords (English)
"Temple", "Travel", "Coffee", "food", "Travel", "Green tea", "Cave", "Forest", "Waterfall", "Tradition", "Places", "Market", "Review", "Atmosphere", "Temple", "Nature", "Travel", "Dam", "Huay", "Travel", "Village", "pagoda", "Cafe", "River", "Mountain", "Festival", "History"," Buddha statues, "Tourists", "Tourist Places", "Local People", "Culture", "Floating Markets", "Festive Places", "Ancient", "Relics", "Attractions", "Museums", "Weekend Markets", "View Points" ","Archaeological site"

This article uses BiLSTM in Text classification of sentiment analysis to improve the result. BiLSTM can be summarized by concatenating the forward and backward states as $ht = [- \rightarrow h\ t, \leftarrow - h\ t]$. It treats all inputs equally. For sentiment analysis, the sentiment contradiction of the text primarily relies on the terms with sentiment data.

In this paper, using BiLSTM can specific group tourism categories from 43 remain to 33 categories with 1,603 comments representing attraction places, and tourism entrepreneurs can use this information to upgrade attraction places.

3.3 Summarizing Results

The results of sentiment analysis and using BiLSTM are shown in Fig. 4. This chart presents that travelers commented positively on the top four places of interest categories: temples (353), food (335), coffee (154), and green tea (142), respectively. The temples praised the beauty of the architecture, the refinement, and the peaceful atmosphere of the village. As for the food, most tourists will comment on its deliciousness. Both temple and food are considered to be highly positive compared to other keywords. As for coffee and green tea, tourists are expected to like the taste and atmosphere of the cafes where they visit, with the number of restaurants growing in popularity, resulting in many positive reviews. Therefore, as a result, if entrepreneurs focus on these four keywords, they can be used to develop tourist attractions, such as developing temples into essential tourist attractions or developing restaurants that focus on delicious taste, good coffee, and good green tea. Or a good atmosphere in the restaurants can increase customers to travel there. Moreover, the graph also shows that unique markets are essential, so perhaps pulling that market up into a niche market can further increase tourism.

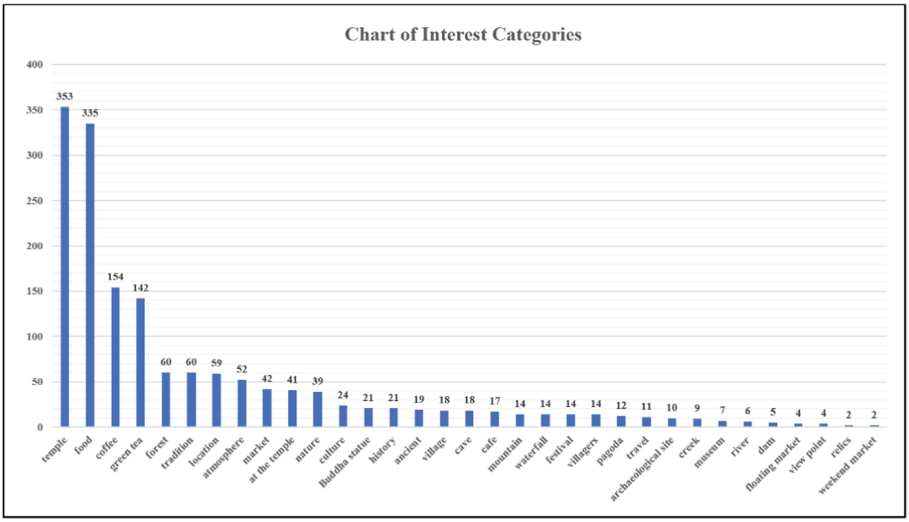

Fig. 4. Tourism categories using BiLSTM

4 Evaluation

In this research, the effectiveness of our proposed model is evaluated by extracting four parameters from the confusion matrix: precision, recall, accuracy, and F1 score. The metrics contain four terms: True Positive (T_P), False Positive (F_P), True Negative (T_N), and False Negative (F_N). *Precision* is the ratio of correctly predicted reviews (T_P) to the total number of expected reviews ($T_P + F_P$) in any class, where the class may be positive or negative. *Recall* is the ratio of correctly predicted reviews (T_P) to the total number of actual reviews ($T_P + F_N$) in any class, where the class may be positive or negative. *Accuracy* is the ratio of correctly predicted to the total number of reviews, and *F1 score* represents the harmonic mean of precision and recall. These four metrics are described briefly below:

$$Precision = \frac{T_P}{(T_P + F_P)} \tag{1}$$

$$Recall = \frac{T_P}{(T_P + F_N)} \tag{2}$$

$$Accuracy = \frac{(T_P + T_N)}{(T_P + T_N + F_P + F_N)} \tag{3}$$

$$F1Score = \frac{2 \times (Precision \times Recall)}{(Precision + Recall)} \tag{4}$$

All the performances of our proposed approach are shown in Table 3.

Table 3. Performance of the proposed text classification

Method	Precision	Recall	Accuracy	F1 Score
Non-BiLSTM	*78.95%*	*80.29%*	*80.40%*	*79.61%*
BiLSTM	***83.25%***	***87.01%***	***85.78%***	***85.08%***

5 Conclusion

This article analyzes the relationship(s) between social media use and its effect on community-based tourism in Thailand using the Social Media Sensing framework (S-Sense) as sentiment analysis and BiLSTM model to analyze the text classification of the comments. The results show that 67.51% of the 19,391 clean-processed comments were satisfied with those attraction places. Therefore S-Sense and BiLSTM models can be satisfactory considering the perspective of comments about attraction places from 43 to 33 keywords with 1,603 comments to lead to tourism development. Travelers commented positively on the top four places of interest categories: temples (353), food (335), coffee (154), and green tea (142), respectively. Therefore, entrepreneurs can focus on these four categories to develop tourist attractions. The evaluation results show that the proposed sentiment analysis approach has heightened precision, recall, and F1 scores. It can enhance to be valuable with increased accuracy on statements.

In the future, we should collect more comments to improve precision in the analytics. The limitation of this article is that the datasets are all in Thai. It would be interesting to execute a parallel study on other languages, such as English or Chinese.

References

1. Lee, T.H., Jan, F.-H.: Can community-based tourism contribute to sustainable development? Evidence from residents perceptions of the sustainability. Tourism Manag. **70**, 368–380 (2019)
2. Melphon Mayaka, W., Croy, G., Cox, J.: A dimensional approach to community-based tourism: Recognising and differentiating form and context. Ann. Tourism Res. **74**, 177–190 (2019). https://doi.org/10.1016/j.annals.2018.12.002
3. Gu, X., Wu, J., Guo, H., Li, G.: Local tourism cycle and external business cycle. Ann. Tourism Res. **73**, 159–170 (2018)
4. Laparojkit, S., Suttipun, M.: The influence of customer trust and loyalty on repurchase intention of domestic tourism: a case study in Thailand during COVID-19 crisis. The J. Asian Finance, Econ. Bus. **8**(5), 961–969 (2021)
5. de Vries, D.A., Peter, J., de Graaf, H., Nikken, P.: Adolescents' social network site use, peer appearance-related feedback, and body dissatisfaction: testing a mediation model. J. Youth Adolesc. **45**, 211–224 (2016)
6. Saiphoo, A.N., Halevi, L.D., Vahedi, Z.: Social networking site use and self-esteem: a meta-analytic review. Pers. Individ. Differ. **153**, 109639 (2020)
7. Kolokytha, E., Loutrouki, S., Valsamidis, S., Florou, G.: Social media networks as a learning tool. Procedia Econ. Financ. **19**, 287–295 (2015)
8. Havakhor, T., Soror, A.A., Sabherwal, R.: Diffusion of knowledge in social media networks: effects of reputation mechanisms and distribution of knowledge roles. Inform. Syst. J. **28**(1), 104–141 (2018)

9. Lee, I.: Social media analytics for enterprises: Typology, methods, and processes. Bus. Horiz. **61**(2), 199–210 (2018)
10. Fan, W., Gordon, M.D.: The power of social media analytics. Commun. ACM **57**(6), 74–81 (2014)
11. Bengio, Y., LeCun, Y., Hinton, G.: Deep learning for AI. Commun. ACM **64**(7), 58–65 (2021)
12. Haruechaiyasak, C., Kongthon, A., Palingoon, P., Trakultaweekoon, K.: S-sense: a sentiment analysis framework for social media Monitoring Applications. Inform. Technol. J. **14**(1), 11–22 (2018)
13. Xu, G., Meng, Y., Qiu, X., Yu, Z., Wu, X.: Sentiment analysis of comment texts based on BiLSTM. IEEE Access **7**, 51522–51532 (2019)
14. Saggi, M.K., Jain, S.: A survey towards an integration of big data analytics to big insights for value-creation. Inform. Process. Manage. **54**(5), 758–790 (2018)
15. Olivera, P., Danese, S., Jay, N., Natoli, G., Peyrin-Biroulet, L.: Big data in IBD: a look into the future. Nat. Rev. Gastroenterol. Hepatol. **16**, 312–321 (2019)
16. Zakir, J., Seymour, T., Berg, K.: Big data analytics. Issues Inform. Syst. **16**(2), 81–90 (2015)
17. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**, 436–444 (2015)
18. Phiphitphatphaisit, S., Surinta, O.: Deep feature extraction technique based on Conv1D and LSTM network for food image recognition. Eng. Appl. Sci. Res. **48**(5), 581–592 (2021)
19. Stieglitz, S., Mirbabaie, M., Ross, B., Neuberger, C.: Social media analytics – Challenges in topic discovery, data collection, and data preparation. Int. J. Inform. Manag. **39**, 156–168 (2018)
20. Holsapple, C., Hsiao, S.-H., Pakath, R.: Business social media analytics: Definition, benefits, and challenges. In: Twentieth Americas Conference on Information Systems (AMCIS), pp. 1–12. Savannah (2014).
21. Khruahong, S., Asawasakulson, A., Krom, W.N.: Social media analytics in comments of multiple vehicle brands on social networking sites in Thailand. In: Luo, Y. (ed.) Cooperative Design, Visualization, and Engineering (CDVE). Lecture Notes in Computer Science, vol. 12341, pp. 357–367. Springer, Cham (2020)
22. Andryani, R., Negara, E.S., Triadi, D.: Social media analytics: data utilization of social media for research. J. Inform. Syst. Informatics **1**(2), 193–205 (2019)
23. Brooker, P., Barnett, J., Cribbin, T.: Doing social media analytics. Big Data Soc. **3**, 205395171665806 (2016)
24. Feldman, R.: Techniques and applications for sentiment analysis. Commun. ACM **56**(4), 82–89 (2013)
25. Medhat, W., Hassan, A., Korashy, H.: Sentiment analysis algorithms and applications: a survey. Ain Shams Eng. J. **5**(4), 1093–1113 (2014)
26. Liu, B.: Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. Cambridge University Press (2020). https://doi.org/10.1017/9781108639286
27. Darwich, M., Noah, S.A.M., Omar, N.: Minimally-supervised sentiment lexicon induction model: a case study of malay sentiment analysis. In: Phon-Amnuaisuk, S., Ang, SP., Lee, Sy. (eds.) Multi-disciplinary Trends in Artificial Intelligence (MIWAI). Lecture Notes in Computer Science(), vol. 10607, pp. 225–237. Springer, Cham (2017).
28. Siami-Namini, S., Tavakoli, N., Namin, A.S.: The performance of LSTM and BiLSTM in forecasting time series. In: IEEE International Conference on Big Data (Big Data), pp. 3285–3292. IEEE, CA, USA (2019)
29. Maghfour, M., Elouardighi, A.: Standard and dialectal Arabic text classification for sentiment analysis. In: Abdelwahed, E., Bellatreche, L., Golfarelli, M., Méry, D., Ordonez, C. (eds.) Model and Data Engineering (MEDI). Lecture notes in Computer Science(), vol. 11163, pp. 282–291. Springer, Cham (2018)
30. Tapsai, C., Meesad, P., Unger, H.: An overview on the development of Thai natural language processing. Inform. Technol. J. **15**(2), 45–52 (2019)