2023 20th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON) | 979-8-3503-1046-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ECTI-CON58255.2023.10153301

Intent Classification from Online Forums for Phuket Medical Tourism

Nasith Laosen
Faculty of Science and Technology
Phuket Rajabhat University
Phuket, Thailand
nasith.l@pkru.ac.th

Kanjana Laosen
Andaman Intelligent Tourism and Service Informatics Center
College of Computing
Prince of Songkla University Phuket Campus
Phuket, Thailand
kanjana.l@phuket.psu.ac.th

Jaturawit Ardharn
Faculty of Radiological Technology
Rangsit University
Pathumthani, Thailand
jaturawit.ar@gmail.com

***Abstract*—Social media makes healthcare and medical information readily available to medical tourists. The medical tourists use social media for searching and communicating about their intents. As the questions posted on social media are rapidly increased, the difficulty to read all questions by human is increased as well. Hospitals running medical tourism business also need to know the needs of medical tourists for improving services and providing the right products to them. The needs or intents of medical tourists can be found on questions that they ask. Therefore, the objective of this study is to collect and classify intents of medical tourists from the questions posted on online forums. In this study, we collect questions related to medical tourism from the TripAdvisor website. We use natural language processing (NLP) to pre-process the questions and classify them using two neural network models, i.e., a BiLSTM model and a BERT model. The experimental result shows that the BERT model provides better performance with 94.22% of accuracy. We also analyze the results and summarize shortcomings of the dataset and the models.**

***Index Terms*—medical tourism, intent classification, BERT, BiLSTM, natural language processing.**

I. Introduction

Thailand becomes one of Asia's leading countries in medical tourism due to the quality of treatment, affordable treatments and the beauty of nature. The healthcare industry tends to grow steadily and become one of important industries of Thailand. As the main components of medical tourism are medical and tourism, Phuket is therefore one of destinations that medical tourists want to visit since it has several hospitals that got the JCI standard (Joint Commission International Standards) as well as a variety of cultures and the beauty of natural resources.

Medical tourists usually seek information about treatment and healthcare through several channels such as search engines, asking the questions through hospital websites, or posting questions on online travel forums like TripAdvisor. The questions posted on those forums indicate tourists' interest (or intent) which is useful for the operation of the medical tourism business.

979-8-3503-1046-7/23/$31.00 ©2023 IEEE

Knowing and understanding the intents of tourists is important for marketing planning and improving quality of service in healthcare industry. Therefore, in this paper, we propose a method for intent classification in the medical tourism domain. In this work, we first crawl data from the TripAdvisor website and create a medical question dataset. Then we construct two neural network models, using bi-directional long short-term memory (BiLSTM) and bidirectional encoder representations from transformers (BERT), for classifying intents. Lastly, based on the classification results, we conduct error analysis to examine shortcomings of the models as well as the dataset.

The rest of this paper is organized as follows: Section II reviews the literature. Section III describes creation of the dataset. Section IV presents experiments and results. Section V analyzes errors. Section VI provides conclusions.

II. Literature Review

A. Neural Network Architectures for Intent Classification

Intent classification is a task in natural language processing (NLP) that involves categorizing a text into a predefined intent category. In particular, intent classification in the medical tourism domain is a task that categorize a text into a medical tourism intent category such as dental care, surgery, cosmetic surgery, and fertility treatments.

There are several neural network architectures that have been used for the intent classification task, e.g., long shot-term memory (LSTM) [1] [2], bi-directional long short-term memory (BiLSTM) [3], bidirectional encoder representations from transformers (BERT) [4] [5] [6], subword semantic hashing [7], and capsule neural networks [8] [9]. Among these architectures, LSTM, BiLSTM and BERT are the most well-known and have been proven to achieve state-of-the-art results. In addition, there exist research works that proposed methods to improve the performance of intent classification by combining models. Some examples of such combination are (1) convolutional neural networks (CNN) combined with the BERT model [10] and (2) CNN combined with the LSTM attention model [11] [12] [13].

BiLSTM is a type of recurrent neural network (RNN). An advantage of BiLSTM is that it is capable of processing

Authorized licensed use limited to: Prince of Songkla University provided by UniNet. Downloaded on September 15,2023 at 04:02:19 UTC from IEEE Xplore. Restrictions apply.

the sequence of an input text in both forward and backward directions. This ability allows BiLSTM to better capture information from both past and future elements of the sequence, which will be helpful in finding better feature representation from the input text [14].

BERT is a language model developed by Google [15]. Basically, a language model represents statistical relationships of words in a language and determines the likelihood of a sequence of words in such language. BERT also consists of an encoder layer built on the Transformer [16] and can be fine-tuned with an additional output layer to create a downstream task such as question answering or text classification. There are several variations of BERT. One of the most popular models is the English uncased BERT-base model, which consists of 12 hidden layers, 768-dimensional hidden layers, 12 attention heads, and 110 million parameters. It was pre-trained using BookCorpus (800 million words) and English Wikipedia (2,500 million words).

B. Word Embedding Techniques

A word embedding technique is a technique for representing words as vectors, where similar or related words are located close to each other in the vector space. This vectors are called word embeddings. Since neural network models cannot process textual data directly, we can create and use word embeddings as input of the models instead. There are several techniques for creating word embeddings. The most widely used techniques are Word2Vec [17] and GloVe [18].

III. Dataset Construction

A. Data Acquisition

On the Tripadvisor website, users can post questions about medical treatments. Therefore, we collected medical questions from the TripAdvisor website under the forum of Phuket province using the WebHarvy tool.[1] We collected the data during April-May 2022 and obtained totally 1,229 questions.

B. Data Pre-processing

We use the Natural Language Toolkit (NLTK) and Keras libraries to pre-process the questions obtained from the previous step as follows:

- Expanding contractions. (For example, the word "I'm" is expanded to "I am".)
- Removing symbols and punctuations.
- Removing all stop words using PorterStemming and lower-casing.
- Tokenizing sentence using the Keras tokenizer.

C. Data Labeling

After pre-processing the data, we count frequency of words in all questions and select only top six words: "Covid", "Dental", "Hospital review", "Insurance", "Surgery", and "Treatment". These six words are corresponding to behavioral medical tourism keywords in [19]. We use these six words as classes of intents (denoted by COVID, DENTAL, HOSPITALREVIEW, INSURANCE, SURGERY, and TREATMENT, respectively). Annotators then categorize each question into a class based on keywords found and overall sentence meaning. The number of text of each class after labelling are shown in Table I and examples of labeled texts are shown in Table II.

TABLE I
THE NUMBER OF QUESTIONS OF EACH CLASS.

Classes (Intents)	The number of questions
COVID	231
TREATMENT	228
DENTAL	210
HOSPITALREVIEW	193
INSURANCE	188
SURGERY	179

IV. Experiments and Results

A. Architecture of Classification Models

As mention in Section I, we create and compare two neural network models, i.e., BiLSTM and BERT models. The details of these two models are described below.

1) The BiLSTM model: An overall architecture of our BiLSTM model is shown in Fig.1. In our experiment, each input text is tokenized and padded to the maximum length of text in the dataset by using the NLTK library. Each token is then replaced by its embedding vector (obtained from applying the GloVe technique). Padding and out-of-vocabulary (OOV) tokens are represented by all-zero vectors at the end. This process helps us to handle sentences with different lengths. The input sequence is then fed through a BiLSTM model. To create the intent prediction, the encoding is multiplied at each time step by a trainable parameter matrix with bias. The result of the linear transform was then flattened. The ReLU activation function is used at the dense layer. The Softmax function is used at the output layer to create a vector of the intent probabilities (6 classes).

2) The BERT model: An overview of our BERT architecture is shown in Fig.2. The model architecture of BERT is multi-layer bidirectional Transformer encoder based on the original Transformer model. In our experiments, we use the English uncased BERT-base model. We fine-tune the model end-to-end by minimizing the cross-entropy loss on our dataset. The input text is lower-cased before tokenized into word pieces. The first token of the sequence is "[CLS]". When there is a pair of sentence, a special token "[SEP]" is added as the final token to separate between sentences.

B. Experiment Settings

The BiLSTM and BERT models are implemented using the Tensorflow library. Both models are trained on 70% of the data, validated on 10% and tested on the remaining 20%. In the training process, we use categorical cross entropy loss, the Adam optimizer, and set a learning rate to 1e-5. Both models are trained using a batch size of 32 for 20 epochs.

[1] https://www.webharvy.com/

Authorized licensed use limited to: Prince of Songkla University provided by UniNet. Downloaded on September 15,2023 at 04:02:19 UTC from IEEE Xplore. Restrictions apply.

TABLE II
THE EXAMPLE OF LABELED QUESTIONS.

Medical Questions	Intent Label
What happens when one of a family gets COVID at phuket internatinal airport?	COVID
Does anyone know of a good dentist in Phuket area?	DENTAL
Bangkok international hospital in Phuket Town. But it is more expensive.	HOSPITALREVIEW
Covered insurance is at Bangkok Phuket Hospital?	INSURANCE
Any one had plastic surgery in Phuket?	SURGERY
Is there a dialysis center on Patong beach?	TREATMENT

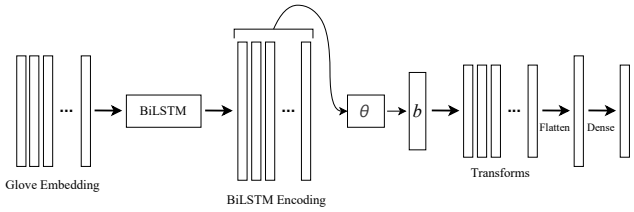

Fig. 1. An overview architecture of our BiLSTM model.

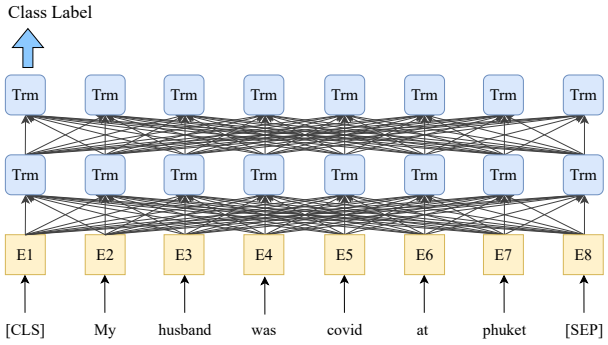

Fig. 2. An overview architecture of our BERT model.

TABLE III
COMPARISON OF ACCURACY OF BILSTM AND BERT MODELS.

Model	Accuracy (%)
BiLSTM	80.10
BERT	94.22

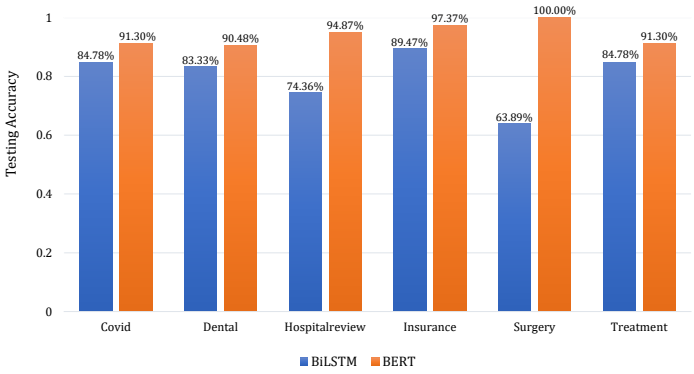

Fig. 3. The accuracy of BiLSTM and BERT models classified by classes.

C. Experimental Results

We compare the effectiveness of the BiLSTM and BERT models on our collected dataset. The results of both models are shown in Table III. As can be seen from the table, the BERT model gives better results with 94.22% of accuracy, while the BiLSTM model gives 80.10% of accuracy. Fig. 3 shows the accuracy of both models classified by classes.

V. ERROR ANALYSIS

We conducted error analysis to better understand pros and cons of the models and the dataset for further improvement. Some interesting issues are listed below.

A. The Size of the Dataset

From the experimental results, we found that the BiLSTM model outputs wrong classification mostly on the class SURGERY (cf. Fig. 3). This error might be due to the dataset being too small and has not enough vocabularies about surgery. In contrast, the BERT model gives the accuracy of 100% on the class SURGERY. This might be due to the fact that the BERT model is a pre-trained model that was earlier trained over unlabeled large corpus. Therefore, the BERT model might has more implicit knowledge about surgery and be able to give higher accuracy.

B. Confusion over Vocabularies

We found that each class has it own key vocabularies. For example, the class DENTAL concerns dental vocabularies such as dental, teeth, veneer, root canal, and other dental services, while the class HOSPITALREVIEW concerns vocabularies about doctors, hospital service, hospital accommodation, or other information related hospitals. However, when these two kinds of vocabularies occur in the same text, it might makes the models confuse, as shown in Example 1 below.

Example 1:
Question: "My husband needs some extensive dental work and it sound like Dr. Joy at Patong Merlin Hotel is highly recommended by many."
Prediction: HOSPITALREVIEW
GroundTruth: DENTAL

In the same manner, Example 2 shows another false classification that misunderstand between the classes TREATMENT and INSURANCE.

Authorized licensed use limited to: Prince of Songkla University provided by UniNet. Downloaded on September 15,2023 at 04:02:19 UTC from IEEE Xplore. Restrictions apply.

Example 2:
Question: "I got an accident and went to Patong hospital. My insurance covered all treatments."
Prediction: TREATMENT
GroundTruth: INSURANCE

In addition, when the models process some unknown vocabularies, they might output wrong classification result as well. Example 3 shows that the model cannot understand medical abbreviations. The word "ENT" stands for "ear, nose, and throat". This cause from having less abbreviation words in the dataset.

Example 3:
Qusestion: "ENT Clinic I would go to Bangkok hospital Phuket."
Prediction: DENTAL
GroundTruth: TREATMENT

VI. CONCLUSIONS

In this study, by crawling medical questions posted on the TripAdvisor website, an intent classification dataset in the medical tourism domain has been created. Two neural network models, i.e., BiLSTM and BERT models, have been built for conducting intent classification over the created dataset. The experimental results show that the BERT model outperformed the BiLSTM model with 94.2% of accuracy. Error analysis has been conducted to identify two shortcomings of the dataset and the models, i.e., the size of the dataset and confusion over vocabularies. From the analysis, we also found that one question can be categorized into more than one class (multi-classes). These challenges lead us to improve the quality of our dataset and models in the future.

ACKNOWLEDGMENT

This study is supported by Super AI Engineer season II program and Andaman Intelligent Tourism and Service Informatics Center, College of Computing, Prince of Songkhla University (Phuket campus), Thailand.

REFERENCES

[1] M. Jbene, S. Tigani, R. Saadane, and A. Chehri, "An LSTM-based intent detector for conversational recommender systems," in *2022 IEEE 95th Vehicular Technology Conference (VTC2022-Spring)*, 2022, pp. 1–5.

[2] G. D. Gennaro, A. Buonanno, A. D. Girolamo, A. Ospedale, and F. A. N. Palmieri, "Intent classification in question-answering using LSTM architectures," in *Progresses in Artificial Intelligence and Neural Systems*. Springer Singapore, jul 2020, pp. 115–124.

[3] E. Haihong, P. Niu, Z. Chen, and M. Song, "A novel bi-directional interrelated model for joint intent detection and slot filling," in *Annual Meeting of the Association for Computational Linguistics*, 2019.

[4] L. Qin, W. Che, Y. Li, H. Wen, and T. Liu, "A stack-propagation framework with token-level intent detection for spoken language understanding," in *Conference on Empirical Methods in Natural Language Processing*, 2019.

[5] Q. Chen, Z. Zhuo, and W. Wang, "BERT for joint intent classification and slot filling," *arXiv preprint arXiv:1902.10909*, 2019.

[6] M. Huggins, S. Alghowinem, S. Jeong, P. Colon-Hernandez, C. Breazeal, and H. W. Park, "Practical guidelines for intent recognition: BERT with minimal training data evaluated in real-world HRI application," in *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 341–350. [Online]. Available: https://doi.org/10.1145/3434073.3444671

[7] K. Shridhar, A. Dash, A. Sahu, G. G. Pihlgren, P. Alonso, V. Pondenkandath, G. Kovács, F. Simistira, and M. Liwicki, "Subword semantic hashing for intent classification on small datasets," in *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2019, pp. 1–6.

[8] C. Xia, C. Zhang, X. Yan, Y. Chang, and P. S. Yu, "Zero-shot user intent detection via capsule neural networks," *arXiv preprint arXiv:1809.00385*, 2018.

[9] C. Zhang, Y. Li, N. Du, W. Fan, and P. S. Yu, "Joint slot filling and intent detection via capsule neural networks," *arXiv preprint arXiv:1812.09471*, 2018.

[10] C. He, S. Chen, S. Huang, J. Zhang, and X. Song, "Using convolutional neural network with BERT for intent determination," in *2019 International Conference on Asian Language Processing (IALP)*, 2019, pp. 65–70.

[11] R. Cai, B. Zhu, L. Ji, T. Hao, J. Yan, and W. Liu, "An CNN-LSTM attention approach to understanding user query intent from online health communities," in *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, 2017, pp. 430–437.

[12] A. S. Varghese, S. Sarang, V. Yadav, B. Karotra, and N. Gandhi, "Bidirectional LSTM joint model for intent classification and named entity recognition in natural language understanding," in *Intelligent Systems Design and Applications*, A. Abraham, A. K. Cherukuri, P. Melin, and N. Gandhi, Eds. Cham: Springer International Publishing, 2020, pp. 58–68.

[13] C. X. Li, C. Zhang, and Q. Fu, "Research on CNN+LSTM user intention classification based on multi-granularity features of texts," *The Journal of Engineering*, vol. 2020, pp. 486–490, 2020.

[14] K. Sreelakshmi, P. C. Rafeeque, S. Sreetha, and E. S. Gayathri, "Deep bi-directional LSTM network for query intent detection," *Procedia Computer Science*, vol. 143, pp. 939–946, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050918320374

[15] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2019. [Online]. Available: https://arxiv.org/abs/1810.04805

[16] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[17] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 2013. [Online]. Available: https://arxiv.org/abs/1301.3781

[18] J. Pennington, R. Socher, and C. Manning, "GloVe: Global vectors for word representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1532–1543. [Online]. Available: https://aclanthology.org/D14-1162

[19] N. Weerakit, C. Chaiyasain, P. Aksornpairoj, and S. Sonpiam, "The development guidelines of health tourism in Phuket to become a medical hub," *Journal of International and Thai Tourism*, vol. 15, no. 2, pp. 35–57, 2020.

Authorized licensed use limited to: Prince of Songkla University provided by UniNet. Downloaded on September 15,2023 at 04:02:19 UTC from IEEE Xplore. Restrictions apply.