IMPLEMENTING A CHATBOT WITH NATURAL LANGUAGE UNDERSTANDING THROUGH BERT AND NEURAL NETWORK INTEGRATION

Wuttipat Nilsiri[1], Ekachai Phaisangittisagul [2*] and Thepchai Supnithi [3*]

[1]Artificial Intelligence and Internet of Thing Engineering, Kasetsart University, THAILAND; Email:wuttipat.ni@ku.th

[2]Department of Electrical Engineering, Faculty of Engineering, Kasetsart University, THAILAND; Email: fengecp@ku.ac.th

[3]NECTEC, National Science and Technology Development Agency, THAILAND; Email: thepchai@nectec.or.th

ARTICLE HISTORY

Received: 3 May 2024 **Revised:** 10 May 2024 **Accepted:** 31 May 2024

ABSTRACT

The objective of this study is to deploy a chatbot tailored for specific domains with a capacity for natural language comprehension. While prior research has demonstrated successful implementations, relying predominantly on Seq2Seq models emphasizing response generation over question comprehension, achieving true natural language understanding necessitates the integration of two core language processing models: intent classification and named entity recognition (NER). However, traditional vectorization methods suffer from a notable drawback in their inability to accommodate unseen words. To address this limitation, we propose a novel approach leveraging the fusion of BERT (Bidirectional Encoder Representations from Transformers) and neural network models. BERT's capacity to capture subword units, synonymous expressions, and inherent word properties as similarity vectors enables robust support for out-of-vocabulary scenarios encountered during language processing tasks. In our experimentation, we utilize the NECTEC sightseeing dataset, preprocessed using BERT embeddings, to evaluate the performance of various neural network models across intent classification and NER tasks. Our findings underscore the promising efficacy of the proposed methodology in enhancing accuracy metrics.

Keyword: Chatbot, NLP, Intent Classification, Named Entity Recognition, Neural Network

CITATION INFORMATION: Nilsiri, W., Phaisangittisagul, E. & Supnithi, T. (2024). Implementing A Chatbot with Natural Language Understanding Through Bert and Neural Network Integration. *Procedia of Multidisciplinary Research*, 2(5),62.

INTRODUCTION

In recent years, the e-commerce sector has experienced rapid expansion, particularly exacerbated by the global Covid-19 pandemic, prompting a widespread transition of brick-and-mortar establishments to online platforms. To facilitate customer communication, businesses typically offer contact channels such as phone calls and live chat functionalities. However, the inherent limitations of these channels pose challenges for effective customer engagement, particularly regarding the transmission of visual information through phone calls and the potential delays associated with live chat interactions. As customer inquiries accumulate, response times may become prohibitively lengthy, despite potential solutions such as increasing call center capacities, which incur additional operational costs. Consequently, this study proposes the development of an AI chatbot leveraging neural network architectures, tailored specifically for the e-commerce domain. While existing chatbot implementations predominantly rely on rule-based systems(Yusril 2018), they often falter when confronted with queries that fall outside predefined rules , resulting in inadequate responses. Prior research efforts have explored Seq2Seq models (Admin 2020; Raj 2021; Giang 2016; Verma et al. 2020), utilizing word2vec for feature representation, yet these approaches primarily focus on converting sentences without achieving true understanding(Ali 2020; Yusril 2018). Although word2vec has demonstrated efficacy in various studies, it lacks the capacity to represent words at a granular level and is unable to handle unknown words effectively (Chunhu Li 2020; Jason P.C. Chiu 2016). Consequently, BERT (Bidirectional Encoder Representations from Transformers) emerges as a promising alternative, offering the ability to encode word properties into vectors and providing robust support for out-of-vocabulary (OOV) words (Jacob Devlin 2019; Chunhu Li 2020; Jetze Schuurmans 2019).

Research Objectives

This study introduces a novel approach to developing a chatbot tailored for close-domain business applications, employing a combination of Bert and Neural Network architectures for intent classification and named entity recognition (NER). This integrated framework facilitates the generation of shared representations capturing both the semantic content and contextual properties of previously unseen words. In close-domain business contexts, customers may pose inquiries spanning a wide range of topics, each potentially associated with distinct intents. To effectively handle this diversity, the chatbot utilizes intent classification techniques to categorize the underlying intent of each query. Additionally, customer queries often encompass multiple pieces of information, serving as query parameters for retrieving relevant responses from a knowledge base or database. Leveraging named entity recognition capabilities, the chatbot can discern and extract pertinent information from user queries, enabling precise query formulation and retrieval of relevant information from the knowledge base for dissemination to the user.

RESEARCH METHODOLOGY

This study presents a novel approach to chatbot implementation tailored for close-domain applications, featuring two primary models. The first model is dedicated to classifying customer intents, facilitating the identification of the underlying purpose behind each customer query. Meanwhile, the second model is designed to extract key information from customer queries to enable accurate responses even when essential details are omitted. Utilizing intent classification, disparate queries sharing a common intent can be mapped to a singular, specific response, contrasting with the Seq2Seq model where each question is mapped to a distinct answer. Furthermore, the approach to information extraction obviates the need for generating redundant query variations based on specific entity words, thereby reducing the dataset requirements compared to the Seq2Seq model. Both models are implemented using deep neural network architectures, integrating Convolutional Neural

Networks (CNN) and Long Short-Term Memory (LSTM) units, with BERT serving as the word embedder to handle words not present in the chatbot's dictionary. The proposed methodology is depicted schematically in Figure 1

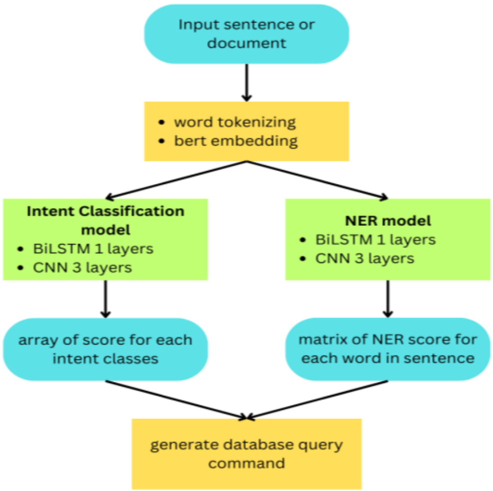

Figure 1 Proposed chatbot implementation

The proposed chatbot architecture, illustrated in Figure 1, comprises two primary components: intent classification and named entity recognition. The Generate Database Query Command module utilizes the outcomes of intent classification and named entity recognition processes, incorporating extracted entities and leveraging the knowledge base, to formulate database query commands. Subsequently, the chatbot generates meaningful responses based on the retrieved information, facilitating effective communication with the user.

(1) Word Tokenizing is a process that separates each word in a sentence into individual words. For Example, our system receives the sentence of อยากไปเที่ยวที่ปทุมธานี, we will separate them into อยาก|ไป|เที่ยว|ที่|ปทุมธานี.

(2) BERT embedding is illustrated in Figure 2.

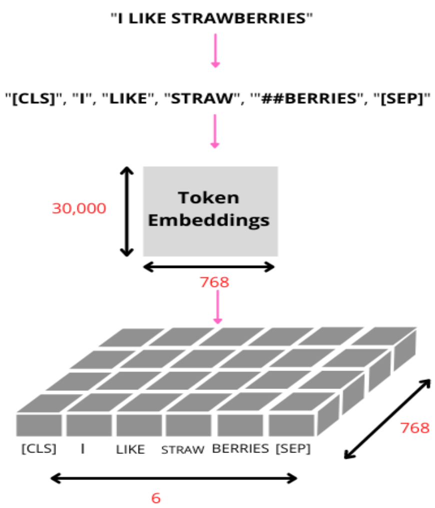

Figure 2. BERT embedding

(2.1) The first token of every sequence of input is always a[CLS], which means the classification token.
(2.2) The final hidden state corresponding to this token is used as the aggregate sequence representation for classification tasks.
(2.3) If the Sentence pairs are packed together into a single sequence, they will differentiate the sentence in two ways. First, using [SEP] token. Second, add a learned embedding to every token indicating whether it belongs to sentence A or sentence B
The token embedding mechanism possesses the capability to discern synonymous terms and to capture inherent word properties, thereby generating similarity vectors. This capability is attained through training on two distinct tasks: Masked Language Modeling (MLM) and Next Sentence Prediction (NSP).

Masked Language Modeling (MLM) constitutes a pre-training methodology aimed at predicting obscured words within a sentence, leveraging contextual cues inherent within the surrounding text.

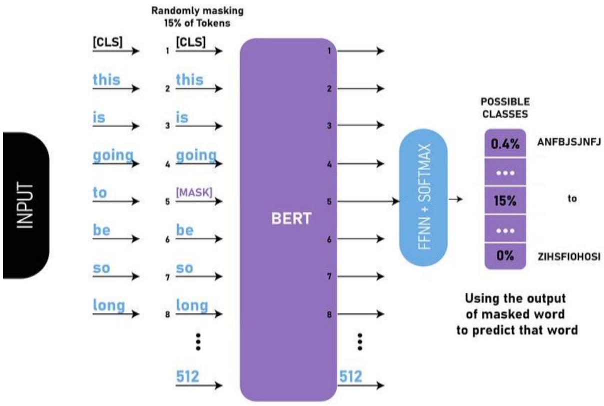

Figure 3. Masked Language Modeling (Jacob Devlin 2019)

Next Sentence Prediction (NSP) represents a task employed in the pre-training of expansive language models, facilitating comprehension of the inter-sentential relationships. The task entails the model's prediction of whether the second sentence logically succeeds the first sentence or not, thereby fostering an understanding of sequential coherence within text.

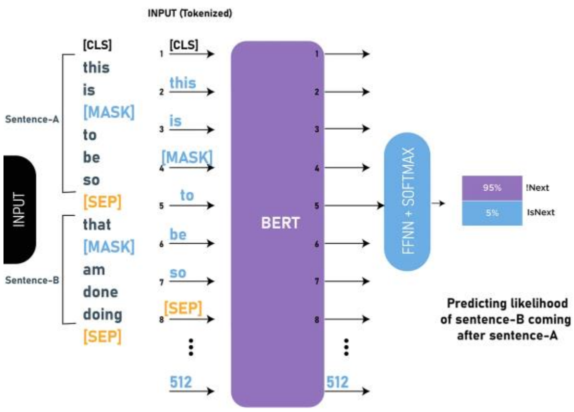

Figure 4. Next Sentence Prediction (Jacob Devlin 2019)

Due to its training on the aforementioned tasks, BERT possesses the capability to generate context-based word vectors, thereby endowing the model with support for out-of-vocabulary scenarios.

(1) Intent Classification and Named Entity Recognition architectures comprise two primary layers consisting of Bidirectional Long Short-Term Memory (BiLSTM) units and Convolutional Neural Network (CNN) layers. In our experimental setup, both models are configured with a single BiLSTM layer and three CNN layers. Within the Intent Classification framework, the window size for each CNN layer is set to 3, resulting in the reduction of vector dimensions as they pass through successive CNN layers. Conversely, in the Named Entity Recognition model, the window size is set to 0, as the model is tasked with predicting named entities for every word within the sentence.

(2) The Generate Database Query Command module serves the purpose of formulating responses to user queries. Leveraging the predictions generated by the intent classification model and named entity recognition model, this module generates database commands designed to retrieve relevant data from the database, subsequently furnishing the retrieved information to the user.

Dataset

To assess the efficacy of the proposed methodology, we utilized a dataset derived from NECTEC's sightseeing data, comprising 5000 question-answering paragraphs. Within these conversations, the maximum participant count is limited to 5 individuals, with an average token count of 142 tokens per paragraph after tokenization. Each paragraph is accompanied by named entity tags denoting the presence of named entities. Our dataset encompasses 15 distinct classes of intent and 11 categories of named entities, as depicted in Figures 5 and 6.

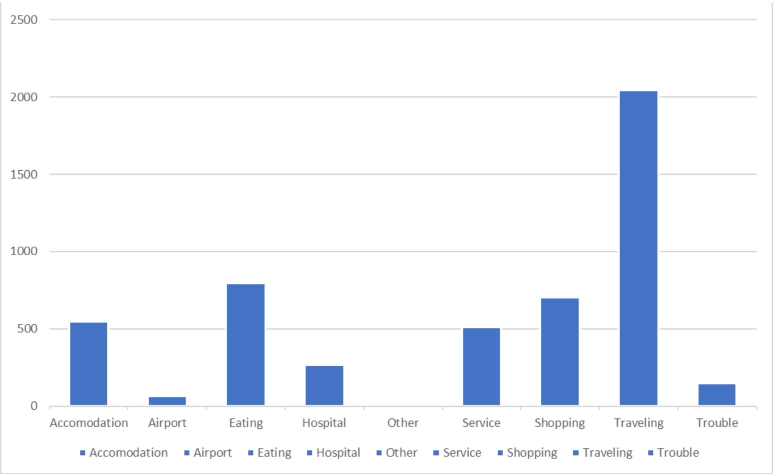

Figure 5. Number of samples each intent class

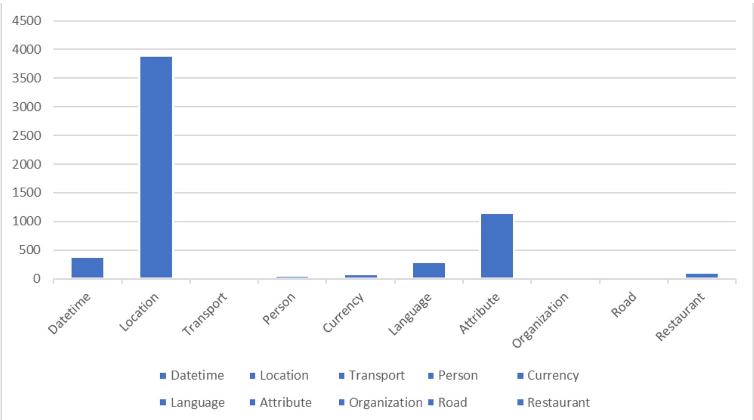

Figure 6. Number of each Named Entity

During the training phase, we employed a stratified 5-fold cross-validation method, especially suitable for classification tasks involving imbalanced data. This approach ensures that each fold utilizes for training and testing maintain a reflection of the class distribution observed within the complete dataset. Additionally, the training regimen comprised 150 epochs, with a batch size configured to 1000.

RESEARCH RESULTS

Intent classification was trained using three distinct models: CNN, BiLSTM, and BiLSTM+CNN, all of which incorporated BERT as the underlying word embedding mechanism. The outcomes obtained from these models are presented in Table 1.

Table 1: Results of stratified 5-fold cross validation of intent classification model

Accuracy	CNN	BiLSTM	BiLSTM+CNN
Fold 1	62%	60%	65%
Fold 2	60%	61%	66%
Fold 3	59%	60%	65%
Fold 4	60%	59%	65%
Fold 5	61%	60%	66%
Average	60%	60%	65%

The experimental results obtained through stratified 5-fold cross-validation reveal that BiLSTM+CNN consistently attains approximately 65% accuracy across all folds on the testing data. This observation suggests that regardless of the dataset partition utilized, the model consistently achieves optimal performance in both training and prediction phases, even when confronted with unseen inputs. Additionally, Figure 7 depicts a confusion matrix contrasting the actual ground truth labels with the labels predicted by the model.

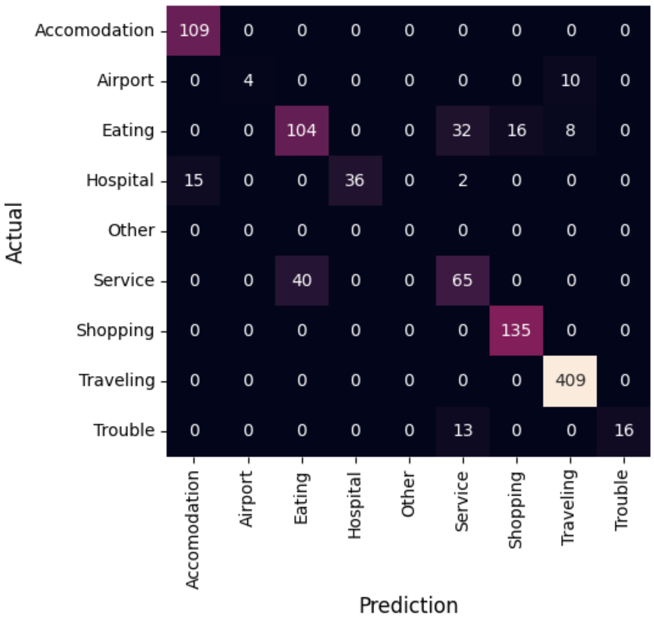

Figure 7. Intent classification confused matrix

For named entity recognition, three models: CNN, BiLSTM, and BiLSTM+CNN were trained utilizing BERT as the word embedding technique. The results of these experiments are presented in Table 2.

Table 2: Results of stratified 5-fold cross validation of Named entity recognition

Accuracy	CNN	BiLSTM	BiLSTM+CNN
Fold 1	84%	80%	86%
Fold 2	82%	79%	85%
Fold 3	80%	81%	84%
Fold 4	82%	79%	85%
Fold 5	82%	81%	85%
Average	82%	80%	85%

The findings from the experimental procedure, conducted using stratified 5-fold cross-validation, demonstrate that BiLSTM+CNN consistently achieves an average accuracy of approximately 85% across all folds when evaluated on the testing dataset, irrespective of the partitioning strategy employed. Furthermore, Figure 8 illustrates a confusion matrix that compares the actual ground truth labels against the labels predicted by the model.

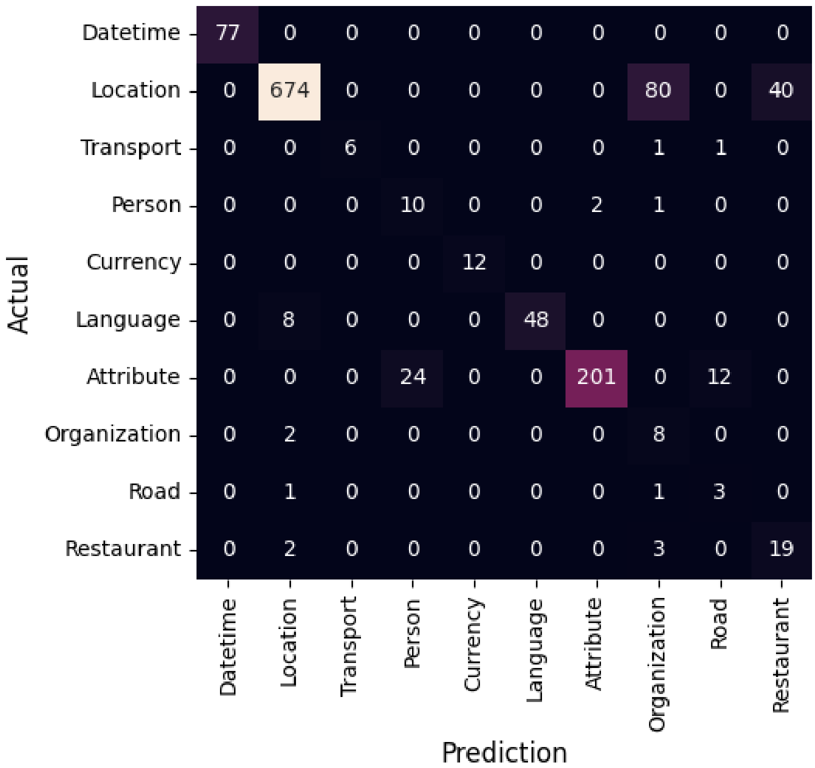

Figure 8. Named entity recognition confused matrix

Tables 3 and 4 present a comparative analysis of performance metrics for intent classification and named entity recognition, respectively, between the proposed method and alternative approaches. It is evident that the proposed method outperforms others in terms of intent classification accuracy and achieves results comparable to the Logan method in named entity recognition.

Table 3: Comparing training performance of the Intent classification model and previous generation methods

Model	Accuracy
Word2Vec and BiLSTM+CNN	52%
Pure BERT	55%
BERT and BiLSTM+CNN	65%

Table 4: Comparing training performance of the Named entity recognition model and previous generation methods

Model	Accuracy
longan	86%
HoogBERTa	80%
BERT and BiLSTM+CNN	85%

CONCLUSION AND DISCUSSION

Based on the experimental findings, our BERT model combined with BiLSTM+CNN architecture achieves an average accuracy of 65% for Intent Classification, surpassing the performance of Word2Vec with BiLSTM+CNN, which only achieves an average accuracy of 52% on the same dataset. This outcome underscores the superiority of BERT over Word2Vec embedding, attributable to BERT's capacity to learn and represent word properties as vectors, thereby providing robust support for out-of-vocabulary (OOV) words. In the realm of Named Entity Recognition (NER), our BERT BiLSTM+CNN model achieves comparable accuracy to HoogBERTa, a model developed by NECTEC employing the BERT architecture for NER tasks. To further enhance performance, potential avenues include the exploration of novel word embedding techniques or the refinement of the inference model structure. Moreover, superior hardware infrastructure is required to effectively train BERT on large datasets, thereby enhancing its familiarity with our specific dataset. In conclusion, this research underscores the utility of BERT as an improvement over Word2Vec for word representation in chatbot applications, offering practical implications for real-world deployment.

ACKNOWLEDGEMENT

The authors are thankful to Assoc.Prof. Kotaro Funakoshi Tokyo Institute of Technology,Japan who provided comments and suggestions that were useful in helping to develop and improve the research even further

REFERENCES

Admin, Jobs. 2020. 'A Simple Introduction to Sequence to Sequence Models'.

Ali, Nazakat. 2020. 'Chatbot: A Conversational Agent employed with Named Entity Recognition Model using Artificial Neural Network'.

Chunhu Li, Chunying Zhang, Qifeng Fu. 2020. 'Research on CNN + LSTM user intention classification based on multi-granularity features of texts'.
Giang, Nguyen Ngoc. 2016. 'DNA Sequence Classification by Convolutional Neural Network'.
Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova. 2019. 'BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding'.
Jason P.C. Chiu, Eric Nichols. 2016. 'Named Entity Recognition with Bidirectional LSTM-CNNs'.
Jetze Schuurmans, Flavius Frasincar. 2019. 'Intent Classification forDialogue Utterances'.
Raj, Vivek Suresh. 2021. 'Performance of Seq2Seq learning Chatbot with Attention layer in Encoder decoder model'.
Verma, Dr. Devraj, Rakshith Yadhav , Sanjana M , Varsha T P , and Vineeth 2020. 'A ChatBot using Seq2Seq and Bag of Words Model'.
Yusril, Muhammad. 2018. 'Comparison Of Multinomial Naive Bayes Algorithm And Logistic Regression For Intent Classification In Chatbot'.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.