

Intent Classification of Users Conversation using BERT for Conversational Dialogue System

Sabyasachi Chakraborty*, Kyung Yul Ohm**, Hyojin Jeon*, Dong Hyun Kim**, Hyoung Jae Jin**

*Terenz Co., Ltd, Busan, Republic of Korea

**TLA Solution Co., Ltd, Seoul, Republic of Korea

saby9996@terenz.ai, ut010135@gmail.com, jeon@terenz.ai, ZS00387@kia-partners.com, ZS00399@kia-partners.com

Corresponding Author: saby9996@terenz.ai

Abstract— Intent Analysis and detection are currently getting a lot of importance for their significance in both the Industry and Academia. The intent classification relies heavily on the rapidly expanding unstructured data of microblogging networks like Twitter and Facebook. However, the work is extremely difficult because of the social media data's frequent noise and diversity. Furthermore, because labelled datasets are usually manually annotated, intent analysis frequently suffers from a lack of data. Modern language representation model BERT (Bidirectional Encoder Representation from Transformers), which accurately models' language, has recently gained interest. In this paper, we developed an Intent Classification Model using BERT for the classification of Questions received from the Users or Humans to specific intents regarding the usage of specific features and components of the car. The dataset for the Classification was developed from the Owner's Manual of the vehicles. The Classification Model was developed using BERT and produced promising results for the classification of Real Work Text into 199 different categories or Intents. The Model which was trained with 254,412 records plotted an overall categorical accuracy of 98.21%
Keywords— Owner's Manual, Car, NLP, NLU, Chatbot, BERT

I. INTRODUCTION

In the past few decades, there has been the continuous improvement of chatbot system research in NLP (Natural Language Processing) and NLU (Natural Language Understanding). On conversational agents, NLU, intent, and entity are extremely valuable ideas manifested in the overall chatbot. NLU in dialogue system stands for the capability that reveals how well chatbots comprehend the natural language of human conversational speech.

Spoken sentences must be separated into words to be classified and identified so that chatbots understand complicated human language. Chatbots take advantage of splendid natural language technology to grasp the grammatical structure of input sentences and classify the user's intent (Park [1]). Particularly, intent classifications in the field of NLP, are one of the most prominent subfields and are the main chatbot conversation engine. Intent classification also analyses text and classifies messages from users into intents that map to corresponding bot actions. In addition, joint intent classification and slot-filling models have been widely studied. And both are NLU operations to classify the intent and detect all relevant slots (entities) for the intent in the query. In general,

intent classification focuses on query intent prediction, whereas slot filling extracts semantic concepts (Chen et al. [2]). If the user says, for instance, "Call a taxi to the school", the chatbot needs to know that the intent of the sentence is "Call a cab" and that destination (i.e., slot) means the entity to call a cab. This enables users to make correct service calls and elevate the experience of using a chatbot.

The paper is organized as follows: The related work and recent studies that have motivated us to perform the work have been depicted in section II. The complete methodology, from data extraction to model development has been discussed in section III. Section 4 further presents the results followed by Section 5 which discusses the paper and lastly the paper is concluded in Section VI.

II. RELATED WORK

Prior studies have investigated tasks related to intent classifications that leverage outstanding Machine Learning algorithms.

Attention-based bidirectional RNN (BiRNN). Liu et al. [3] investigate how an SLU (Spoken Language Understanding) model can profit from the powerful modeling capabilities of the sequence architecture. Specifically, the authors discovered an attention-based bidirectional RNN (BiRNN) structure for joint intent detection and slot filling which are inspired by the usage of an attention mechanism in an encoder-decoder architecture. Attention-based RNN model is computationally more remarkable compared to the attention-based encoder-decoder structures that employ explicitly ordered inputs. While model training, whereas the proposed attention-based RNN process the input sequence only once, the encoder-decoder slot filling model, and the encoder-decoder slot filling model implement the input twice. In their research, the proposed models accomplish cutting-edge intent detection error rates and slot-filling F1 scores on the task of benchmark ATIS.

BERT (Bidirectional Encoder Representations from Transformers). In a joint intent classification and slot-filling model, Chen et al. [2] in their research have suggested a remarkable model based on BERT for natural language understanding. The BERT model they proposed is pre-trained with two tactics on large-scale unlabelled text, namely, masked

language model and next sentence prediction. The pre-trained BERT then indicates a robust context-dependent sentence representation and can be utilized for a variety of target tasks, namely intent classification, and slot-filling, via fine-tuning procedures, that are comparable to how they are applied for other NLP tasks. Their experimental results also present what their findings demonstrate significant improvements in intent classification accuracy, slot filling F1, and sentence-level semantic frame accuracy.

ConvertRT (Conversational Representations from Transformers). However, Henderson et al. [4] explain that pretrained models such as BERT are computationally heavy, slow, and expensive to train, making them unsuitable for real conversational agents. Thus, their research suggests ConveRT, a pre-training framework for conversational tasks that meets all of the following requirements: a new lightweight model of neural response selection for conversation that can be effectively, affordably, and fast trained. In the study, ConveRT also construct an architecture that considers multiple rotational contexts known as Multi-context ConveRT. In addition to efficient pretraining, since only the classification layer is trained for ConveRT, intent classifiers based on ConveRT encoding train 40x faster than those based on BERT.

III. MATERIALS AND METHODS

A. Data Acquisition and Processing

The dataset for the Training of the Intent Classification model for the Natural Language Understanding System of the Owner's Manual Chatbot was created and curated by TLA Solution Co., Ltd. The classes or the labels of the dataset were the intents that can be the destination of a user's question at a given instance. The features on the other side were the questions that can be asked by any user concerning a particular instance. For example, for an intent to get the information of Occupant Detection System

Intent: *occupant_detection_system_(ODS)*, the questions can be **Q1:** *What is ODS*, **Q2:** *Info about Occupant Detections System* or **Q3:** *ODS? Etc.*

Therefore, in total, 199 Unique Intents were created from the Owner's Manual of a specific car. Therefore, for the 199 Intents, a total of 299,672 questions were formed. By keeping an average of 1505 Questions for each intent.

B. Data Pre-processing

The pre-processing of the data is of prime importance for the development of any deep learning-based model. The data processing of the probable questions of the user was divided into particularly three steps:

1. Remove Special Characters and Stop Words
2. Tokenization and Named Entity Recognition
3. Convert the Sequence of Tokens to Numbers and Padding the Data to the Same Length

1) Removing Special Characters and Stop Words:

The removal of the special characters and stop words are of prime importance. As the Intent Classification is specifically

for the owner's manual of the car, therefore, the questions are accepted to be somewhat technical hence the flow of the questions is less important than the objects in the questions. Therefore, special characters such as (Question Marks (?), Commas (,), Exclamation Mark (!), Colon (:), etc were completely removed from all the text.

Apart from Special Characters and Stop Words, specific data points were also removed from the dataset as they were duplicates of one another. Therefore, after the data points removal a total of 254,512 data points remained.

2) Tokenization and Named Entity Recognition

The Tokenization and Named Entity recognition of the Text is of primary importance to feeding the data to the BERT Input layer [5]. BERT Specifically takes sequences of numbers as input for the classification. Hence the conversion of text to a sequence of numbers is of prime importance.

Moreover, the conversion of the text to numbers must be done in a way where the tokenization of a specific named entity in the text should be established with a specific id. Therefore, to satisfy this a Named Entity Recognition Model was developed as the pretrained Named Entity Recognition models do not suffice our needs because the data set on which we train is far different from the dataset on which the pretrained models have been trained. Moreover, the classes that are being represented by the pretrained Name Entity recognizers are all very general and our Named Entities are very specific to components of a Car.

For the Named Entity Recognition, the test sequences were tokenized and each token was labelled with Five (5) types of classes (O, B-Comp, I-Comp, B-Task, I-Task). Table 1 below shows the distribution of each class and also the definition of each class to which a specific token is mapped.

Class	Definition	Number
O	Un Defined	687,122
B-Comp	Start of a Car Component	299,782
I-Comp	Continuation of Car Component	185,423
B-Task	Start of Task Intention	247,189
I-Task	Continuation of Task Intention	98,568

The NER Model was further re-trained on the Bert Token Classification model to build a new tokenizer for our specific use case. Further, the Tokenizer will be used to convert the Training Data to tokens to feed it into the Intent Classification Model.

3) Convert the Sequence of Tokens to Numbers and Padding the Data to the Same Length

The Training data that we have pre-processed in Step 1 now needs to be prepared for feeding it into the Intent Classification Model which has been powered with the BERT. But for feeding the Training Data, the data needs to be converted into a sequence of tokens which will be further substituted as a

number. Therefore, for creating Tokens the Text will be Tokenized and will be converted to the tokens using the Named Entity Recognition Model developed in step 2.

Figure 1 below shows the Tokenized Text, for a single input text. The ids in the Input Tensor more than 0 are text tokens and 0s represent the padding in the text which are added to make all input vectors of the same size.

active air flaps info about

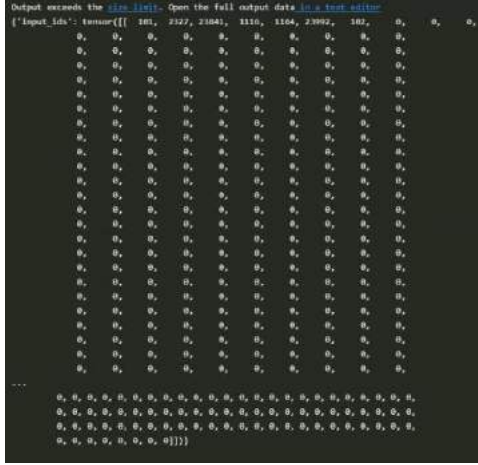


Figure 1. Example of Tokenizer Output

C. Model Development

Recent evolution in the field of machine intelligence has created a very large paradigm shift to solve some of the most precise and difficult issues. But mostly after the emergence of Deep Learning and Artificial Intelligence, the world of Conversational Agents has seen a rapid disruption. Previously, where the conversational agents used to flow in a rule-based manner, now with the advent of Deep Learning and Artificial Intelligence the conversational agents flow with the input conversations made by the user. But the primary component for any Conversational Agent is the method or module to understand the user's input. The User's Input is primarily understood by the use of the Natural Language Understanding Module which is further powered by Intent Classification Models. The Intent Classification Models primarily take the input of the user and match it to a specific intent to understand the question or the input of the user.

For our specific study, we would use BERT (Bidirectional Encoder Representations from Transformers) for the development of an Intent Classification Model. The Bidirectional Encoder Representations from Transformers (BERT) model is a pre-trained language model based on Transformers that was developed by Google Brain Team in 2018 [6]. It outperforms the General Language Understanding Evaluation (GLUE) Benchmark on 11 natural language processing tasks.

Semi-supervised learning was used to train the BERT model using Wikipedia as the training data. Sequence pairs with

masks covering random text words serve as the BERT's training data. The BERT model has two training objectives: predicting masked words and determining whether the two sequences from the input data have context relations. The BERT model is trained to comprehend input text at both the word- and sentence level, enabling it to adapt to different downstream linguistic processing tasks.

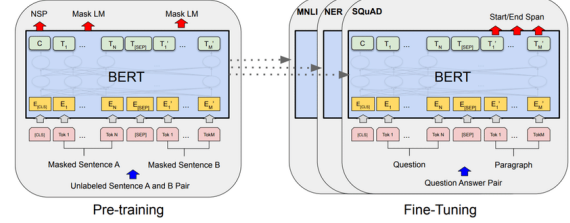


Figure 2. Fine Tuning and Training Process of BERT Model[7]

The BERT model transfer learning process, including pre-training and fine-tuning, is depicted in Figure. 2 in a simple form. The model learns how to combine words and phrases during pre-training because of being trained on a large amount of data. The inheritable parameters for these understandings are stored. The model is retrained for a new task during fine-tuning. Due to the inherited parameters, the model can achieve its optimum state with only a small amount of additional information. [7]

Hence, the tokenized output from the tokenizer is used to provide the input to the BERT Model used in the work.

The Adam (Adaptive Moment Estimation) [8] optimizer was used by the Intent Classification Model established in this work to optimize the cost function, which is the categorical cross-entropy. The optimizer algorithm Adam's hyperparameters were tuned using Bayesian Sequential Model Optimization. Bayesian SMBO is a hyperparameter optimization that reduces a specific objective function by building a surrogate model from the objective function's prior evaluation results [9].

The set of hyperparameters that were chosen for the optimizer function were as follows, learning rate: 0.0001, beta-1: 0.93241, and decay: 0.0000024. Moreover, for the learning rate, a certain policy was defined where whenever for 5 consecutive epochs if the validation loss doesn't decrease then the learning rate must be reduced by a fraction of 0.2.

IV. RESULTS

The Intent Classification model that was created for the study gave us some solid outcomes regarding the accuracy of the classification of user-asked questions to particular intent classes. The created model produced an average f1 score of 0.9801 and categorical accuracy of 98.21% average recall of 0.973 and average precision of 0.985. Additionally, 5-Fold cross-validation was also done on the entire dataset to verify the

generalizability of the model. The model also showed a consistent tendency toward precision and recall over random data folds of training, validation, and testing. There were 25,451 Questions from all 199 Classes in the dataset used to test the model.

Figure 3 described below shows the Average Recall and Loss chart of training the Intent Classification Model. The performance graph shows that the model ran for 100 epochs whereas while fitting the model a maximum limit of the running of the epoch was kept to be 500 while accompanying an early stop protocol. The early stop protocol was devised in such a way that, if the validation loss does not decrease for consecutive 8 epochs, then the training needs to be stopped.

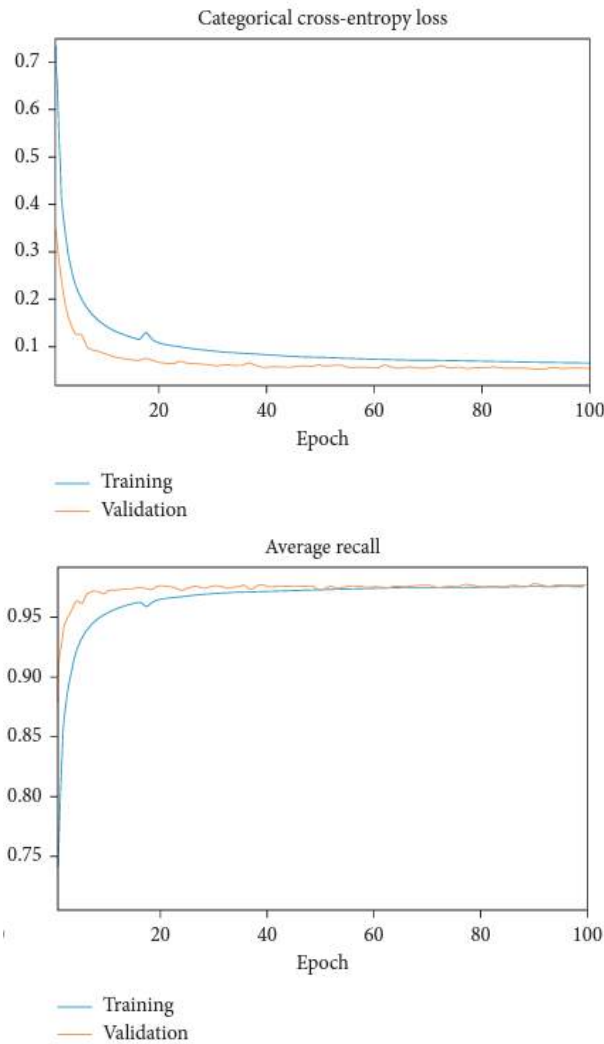


Figure 3. Figure 3 Average Recall and Loss Curve

V. DISCUSSION

The Intent classification Models that have been fine-tuned and trained using the Owner's Manual Data pursue some astounding results in classifying the Questions and Responses of the user to accurate Intents. Such kinds of classification

models are very much necessary for conversational agents as these models unleash the power of Natural Language Understanding. Further these Natural Language Understanding modules can further understand the intent of users even if the inputs contain typos or wrong grammar. Therefore, presently in the world of Non-Face-To-Face Systems and the 4.0 Industrial Revolution, there is a dire need for conversational agents that can provide the users or the Owners of the vehicles in our specific case with 24X7 service.

In the present work, we specifically used the data from the Owners Manuals of Cars which are curated by TLA Solutions. Further with the data, a 5 class Named Entity Recognition Model was developed using Bert Tokenizer which further supported the Tokenization of our input text before feeding it into the model. The Tokenized data was further padded and converted to numerical identities and were fed into the model for training. The Model that was primarily configured for the Intent Classification was a BERT Model. The training data points that were fed into the BERT Model were the Numerical Ids of the Tokenizer Output and the classes were the Intents. Further, the Model was trained for around 100 Epochs with ADAM as an optimizer.

VI. CONCLUSION

In this paper, an Intent Classification Model was discussed for the classification of users' questions to specific intents, when they ask about the parts, issues, and usage of different features of cars. The Intent Classification model was trained with around 254,512 records of questions belonging to 199 different classes or Intents. The Intent Classification Model was developed using BERT (Bidirectional Encoder Representations from Transformers). The present model plotted an overall categorical accuracy of 98.21%.

The methodology provided in the paper can be further improvised by using more data and also data from call centers of Car Companies who face real issues with the customer's queries. The questions or the data that are fetched from the call centers will have the primary entropy and variability which will lead to better generalization of the model. But the present system enabled us with much confidence that this kind of system can be employed in real-life scenarios and stable production environments.

ACKNOWLEDGMENT

This research was supported by the Ministry of Science and ICT, National IT Industry Promotion Agency(NIPA).

REFERENCES

- [1] Park, P.W., 2020. Text-CNN Based Intent Classification Method for Automatic Input of Intent Sentences in Chatbot. *Journal of Korean Institute of Information Technology*, 18(1), pp.19-25.
- [2] Chen, Q., Zhuo, Z. and Wang, W., 2019. Bert for joint intent classification and slot filling. *arXiv preprint arXiv:1902.10909*.
- [3] Liu, B. and Lane, I., 2016. Attention-based recurrent neural network models for joint intent detection and slot filling. *arXiv preprint arXiv:1609.01454*.

- [4] Henderson, M., Casanueva, I., Mrkšić, N., Su, P.H., Wen, T.H. and Vulić, I., 2019. ConveRT: Efficient and accurate conversational representations from transformers. arXiv preprint arXiv:1911.03688.
- [5] Litake, O., Sabane, M., Patil, P., Ranade, A. and Joshi, R., 2022. Mono vs multilingual BERT: A case study in hindi and marathi named entity recognition. arXiv preprint arXiv:2203.12907.
- [6] Purohit, H., Dong, G., Shalin, V., Thirunarayan, K. and Sheth, A., 2015, December. Intent classification of short-text on social media. In 2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity) (pp. 222-228). IEEE.
- [7] Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference. pp. 4171--4186 (2019)
- [8] Frazier, P.I., 2018. A tutorial on Bayesian optimization. arXiv preprint arXiv:1807.02811.
- [9] Chakraborty, S., Aich, S., Kumar, A., Sarkar, S., Sim, J.S. and Kim, H.C., 2020, February. Detection of cancerous tissue in histopathological images using Dual-Channel Residual Convolutional Neural Networks (DCRCNN). In 2020 22nd International Conference on Advanced Communication Technology (ICACT) (pp. 197-202). IEEE.



Sabyasachi Chakraborty is presently working as the Chief Technology Officer of Terenz Co., Ltd, Busan, Republic of Korea. He has previously worked as a Research Scholar at Inje University and worked on the development of Artificial Intelligence Algorithms for Healthcare and Medical Devices. He has wide range of experience in Industry and Academia. His Research Interests are Machine Learning, Artificial Intelligence, Bio Signal Processing and Brain Computer Interface



Hyojin Jeon is presently working as a Senior Researcher, at Terenz Co., Ltd, Busan, Republic of Korea. She has previously worked in vividly in Information Technology Company in Singapore. Her Research Interest is Electronics Engineering, Mechatronics Engineering and Artificial Intelligence.



Kyung Yul Ohm is presently working as a Managing Director at TLA Solution Co., Ltd, Seoul, Republic of Korea. He worked in integrated purchasing and export management for Yura Co., Ltd. and as the head of the general management team of overseas corporations of Sewon Precision Co., Ltd. His Research Interest are Automobile Design, Automobile Engineering, Self-Driving Cars and Artificial Intelligence



Dong Hyun Kim is presently working as a Director at TLA Solution Co., Ltd, Seoul, Republic of Korea. He worked as the head of planning office of Sewon Precision Co., Ltd. and has been in charge of management and new system development at TLA Solution Co., Ltd. His Research Interest are Automobile Design, Automobile Engineering, Self-Driving Cars and Artificial Intelligence



Hyoung Jae Jin is presently working as a Deputy Senior Manager at TLA Solution Co., Ltd, Seoul, Republic of Korea. He worked as maintenance guidelines engineer at Kyungjin Technical Solution and was a part of VR Team at AST Holdings Co., Ltd. His Research Interest are VR, AR, XR, MR, Automobile Engineering, Self-Driving Cars and Artificial Intelligence