Text Intent Classification

Vansh Mahendra, Saloni Reddy, and Shreya Kumari
20BCE7015, 20BCI7167, 20BCI7293

***Abstract* -**
This paper presents an investigation of three different natural language processing (NLP) models for intent classification on the Airline Travel Information System (ATIS) dataset. The ATIS dataset is a widely used benchmark dataset in the field of NLP for its practical relevance in the airline reservation domain. The three models used in this study are Support Vector Machines (SVM), Long Short-Term Memory (LSTM), and Bidirectional Encoder Representations from Transformers (BERT). Our results show that BERT outperformed both SVM and LSTM, achieving an accuracy of 98.5%. This indicates that contextualized word embeddings can effectively capture the complex nuances and variations in the language that is present in the ATIS dataset. The results also highlight the importance of selecting appropriate NLP models for specific tasks and datasets. Our findings have practical implications for the airline industry, as accurate intent classification can lead to improved customer service and operational efficiency. Overall, this study demonstrates the effectiveness of NLP models for intent classification on the ATIS dataset and provides insights into the strengths and limitations of different models.

Index Terms - BERT, Intent Classification, LSTM, SVM

Introduction

Natural Language Processing (NLP) has emerged as a critical area of research in recent years, owing to its ability to automate language processing tasks such as sentiment analysis, machine translation, and intent classification. Intent classification, in particular, is a critical task in the domain of NLP that involves categorizing natural language utterances into pre-defined intent categories. The classification of utterances into multiple classes, also known as multiclass classification, is a common requirement in real-world NLP applications.

In this paper, we present our investigation of three NLP models - Support Vector Machines (SVM), Long Short-Term Memory (LSTM), and Bidirectional Encoder Representations from Transformers (BERT) - for multiclass intent classification on the Airline Travel Information System (ATIS) dataset. The ATIS dataset is a widely used benchmark dataset in the field of NLP, consisting of natural language utterances related to airline travel and categorized into 8 intent categories. SVM is a traditional machine learning algorithm that has been widely used in the field of NLP. It works by mapping data points into a high-dimensional space and identifying a hyperplane that separates the classes. LSTM, on the other hand, is a type of neural network that can capture the sequential nature of language and has been shown to perform well in NLP tasks. BERT is a state-of-the-art transformer-based model that has achieved impressive results in a range of NLP tasks. Our study involves using SVM as the first model to perform basic multiclass classification on the ATIS dataset. We then compare the performance of LSTM and BERT against SVM to evaluate their effectiveness in capturing the nuances of natural language. Multiclass classification is a challenging task, and we aim to identify the most effective model for this task, while also providing insights into the strengths and limitations of each model.

Overall, our study contributes to the ongoing research in the field of NLP and provides insights into the effectiveness of different models for multiclass intent classification. The findings of this study have practical implications for the airline industry, where accurate classification of natural language utterances can lead to improved customer service and operational efficiency.

Related Work

The task of intent classification has received significant attention in the field of NLP, and a range of models have been proposed to address this task. In this section, we review some of the related work in the area of intent classification, with a focus on multiclass classification.

In a 2023 paper, the authors propose a self-supervised method in which they modelled the classifying short text corpus as a heterogeneous graph to address the information sparsity problem and subsequently proposed a heterogeneous graph neural network model to exploit internal and external similarities among short texts. [1]

On reviewing several papers from 2021 and 2022, it was found that basic ml models such as K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), and Gaussian Naive Bayes (GNB) are very widely used for text classification, especially to perform traditional emotion detection from given dialogues.[2][3][4]

Traditional machine learning algorithms, such as SVM and Naive Bayes, have been widely used for intent classification tasks. SVM, in particular, has been shown to be effective in identifying intent categories from natural language utterances. Several studies have also explored the use of neural networks for intent classification, with LSTM being a popular choice. LSTM has been shown to capture the sequential nature of language and has achieved high accuracy in intent classification tasks[5][6].

In recent years, transformer-based models, such as BERT, have emerged as state-of-the-art models for a range of NLP tasks. BERT utilizes a pre-training approach that captures the contextualized representation of words, which has been shown to be effective in capturing the complex nuances of natural language[7]. Several studies have shown the effectiveness of BERT in intent classification tasks. The result suggests that the application of BERT transfer learning and different variations of it has reduced the volume requirement for intent classification tasks to a satiable level.[8][9][10]

Overall, the related work suggests that SVM, LSTM, and BERT are effective models for intent classification tasks, with each model having its strengths and limitations. Furthermore, deep learning models, such as LSTM and BERT, have been shown to be effective in multiclass classification tasks, highlighting the importance of selecting appropriate models for specific tasks and datasets.

Dataset

The Airline Travel Information System (ATIS) dataset is a commonly used benchmark dataset for intent classification tasks in natural language processing. It consists of human-computer dialogues in the context of airline travel information systems, with a focus on identifying the user's intent from their spoken or written input. The dataset contains 4,978 examples of user queries, each labelled with one of 21 intent classes such as flight information, ground service, and airfare information.

ATIS dataset provides a large number of messages and their associated intents that can be used in training a classifier. Understanding the intent of the customer is key to implementing a smooth experience for the end user. The ATIS dataset is widely used for evaluating intent classification models and has been used as a benchmark for various research studies.

	label	query
0	atis_flight	i want to fly from boston at 838 am and arriv...
1	atis_flight	what flights are available from pittsburgh to...
2	atis_flight_time	what is the arrival time in san francisco for...
3	atis_airfare	cheapest airfare from tacoma to orlando
4	atis_airfare	round trip fares from pittsburgh to philadelp...
5	atis_flight	i need a flight tomorrow from columbus to min...
6	atis_aircraft	what kind of aircraft is used on a flight fro...
7	atis_flight	show me the flights from pittsburgh to los an...
8	atis_flight	all flights from boston to washington
9	atis_ground_service	what kind of ground transportation is availab...

Fig. 1 Sample of ATIS Dataset

Methodology

DATA PREPROCESSING

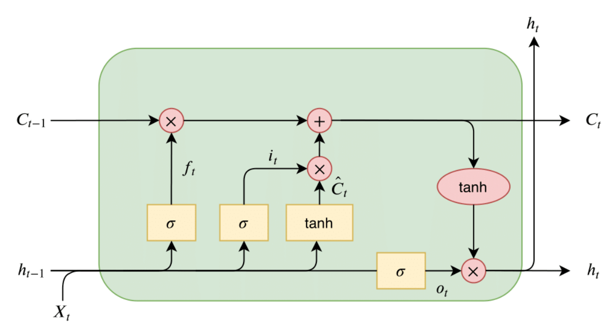
Spacy is an open-source library in Python used for various Natural Language Processing (NLP) tasks such as tokenization, part-of-speech tagging, dependency parsing, named entity recognition, and text classification. Spacy provides a variety of pre-trained language models that can be used to generate vector representations of words and sentences. These pre-trained models are trained on large corpora of text and use advanced techniques such as neural networks to generate high-quality vector representations. The list of input sentences is encoded to a 2D numpy array containing the vector representations of the sentences, each row represents the vector representation of a sentence. This encoding step is a crucial part of various NLP tasks such as text classification, where the vector representations of sentences are used as input features for machine learning models. These models learn to map the input sentence vectors to a set of pre-defined classes, such as the intent categories in an intent classification task. Various text pre-processing techniques such as stopword and punctuation removal, stemming, and lemmatization were performed to clean and analyze the text data, reduce the dimensionality of the data and removal of noise, which ultimately led to better performance.

A padding of 20 was applied to each preprocessed sentence to standardize the length of input sequences. This was done to create a fixed input size for the model and enable the model to process the input sentences in a consistent and standardized way.

MODEL SPECIFICATION

Using the training data and labels, the SVM model is trained, and the resulting model can be used to predict the intent category of new, unseen sentences. Because of its capacity to handle high-dimensional data, such as vector representations of sentences in this example, and its ability to generalise effectively to unknown data, the dataset was first trained on SVM.

The second model includes an input layer, an LSTM layer, two fully connected dense layers, batch normalization, dropout, and softmax activation for the output layer. The model is trained using CategoricalCrossentropy loss, AUC metrics, and Adam optimizer. LSTM is a sequential model with multiple layers. The first layer is the Input layer. It takes in 3D input of shape `(steps, input_dim)`. The second layer is the LSTM layer which processes the input sequence. The third layer is the Dense layer that maps the output of the LSTM layer to the desired output shape. It uses the ReLU activation function for non-linearity. The fourth layer is the BatchNormalization layer which normalizes the output of the dense layer to prevent overfitting. The fifth layer is the Dropout layer which randomly drops out a fraction of the neurons to prevent overfitting. The sixth layer is another Dense layer that maps the output of the previous layer to the desired output shape. It uses the Glorot uniform initializer and a softmax activation function to generate probabilities for each class. The `kernel_initializer`, `bias_initializer`, `kernel_regularizer`, and `bias_regularizer` are regularization techniques that are used to avoid overfitting.



\Fig. 2 LSTM Architecture

CategoricalCrossentropy (CC) is utilised as the loss function, AUC is used as a metric, and Adam is used as the optimizer. CategoricalCrossentropy is a commonly used loss function in multi-class classification issues that computes the cross-entropy loss between predicted probability and true labels. It is ideal for the intent classification problem in this project, where we have numerous classes and want to minimise the discrepancy between projected and actual labels. AUC (Area Under the Receiver Operating Characteristic Curve) is a popular statistic for evaluating binary classification problems. It computes the area under the ROC curve to assess a model's ability to discriminate between positive and negative classes.

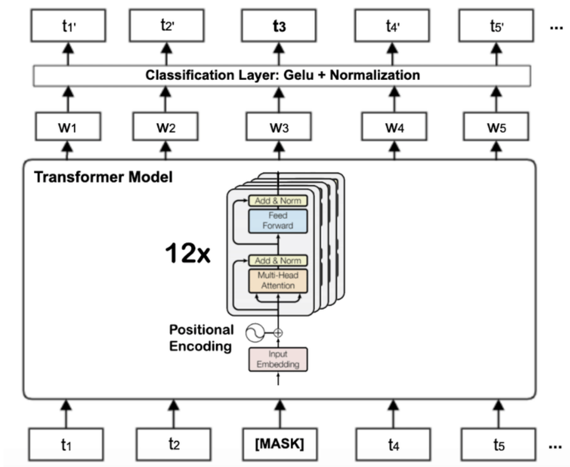


Fig. 3 BERT Architecture

Google's BERT (Bidirectional Encoder Representations from Transformers) is a cutting-edge pre-trained language model. It uses the transformer architecture to generate high-quality embeddings for natural language problems. BERT's architecture is made up of a multi-layer bidirectional transformer encoder. The transformer encoder is made up of many self-attention layers, each of which computes attention scores between all pairs of input embeddings and utilises these scores to weigh the value of each input token in relation to the importance of each other token. The weighted embeddings that arise are then sent into a feedforward neural network to produce the final output. We may use BERT to do intent classification by fine-tuning the pre-trained model on our unique goal by adding a classification layer on top of the pre-trained model. The weights of the pre-trained model are frozen during fine-tuning, and only the weights of the newly added classification layer are updated. The model's input is the sequence of tokens in the input sentence, and its output is a vector expressing the likelihood of each intent class. Here, we have used the BERT model with 8 transformer layers, 512 hidden units per layer, and 8 attention heads per layer, using uncased English text. The model takes a text input, applies a preprocessing layer using a specified TensorFlow Hub handle, encodes the text using a BERT encoder layer with the same specified handle, and then applies a dropout layer followed by a dense layer with 8 output units and no activation function. The raw output is then passed through the softmax activation function which returns a probability distribution over the classes. CategoricalCrossentropy (CC) is utilised as the loss function, Categorical Accuracy is used as a metric, and Adam is used as the optimizer.

Results and Discussion

Our SVM model with decision shape oneVsOne and RBF kernel obtained **96%** accuracy on test data.

The AUC metric is used to evaluate the performance of the LSTM model. The area under the curve (AUC) is a measure of the accuracy of the binary classifier model in distinguishing between positive and negative classes. In this case, AUC is used to evaluate the multiclass classification problem. AUC values range between 0 and 1, with higher values indicating better model performance. LSTM AUC value was 99.59%. The accuracy obtained on the test set was **98%**.

Results show that the BERT model achieved an overall Loss of 6.5% and an Accuracy of **98.5%** on the test set. The training set had a loss of 1.42% and categorical accuracy of 99.77%, while the validation set had a loss of 0.71% and categorical accuracy of 97.07%.

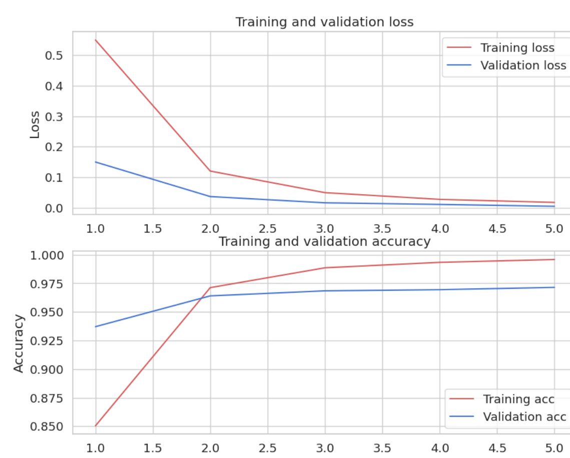


Fig 4. Loss and Accuracy of Training v/s Validation Set

Conclusion and future scope

In this study, we compared the performance of three models, SVM, LSTM, and BERT, for intent classification on the flight dataset. The results show that SVM achieved an accuracy of 96%, LSTM achieved 98% accuracy, and BERT achieved the highest accuracy of 98.5%. Accuracy was chosen as the best measure of model performance since it provides a simple and straightforward way to evaluate model performance. Based on these results, we conclude that BERT is the most suitable model for intent classification on the flight dataset, and can provide valuable insights and solutions for the airline industry.

In conclusion, this study presents an investigation of three different NLP models for intent classification on the ATIS dataset. These results highlight the effectiveness of contextualized word embeddings for capturing the complex nuances and variations in language present in the ATIS dataset. It also emphasizes the importance of selecting appropriate classification models according to the required tasks and datasets. The practical implications of accurate intent classification in the airline industry make this study valuable for improving customer service and operational efficiency.

The study presented in this paper offers potential avenues for future research in natural language processing and intent classification on the ATIS dataset. One of the main expanding features can include investigating other state-of-the-art NLP models that can be explored for intent classification on the ATIS dataset. On the other side of this lies the concept of exploring other domains where intent classification can be useful, such as healthcare, finance, and education. However, the most feasible future development that can be implemented on this project is the development of a real-time intent classification system that utilises the ability to accurately classify user intents in real-time can be valuable for providing timely and efficient customer service. Future research could explore the development of a real-time intent classification system for the airline industry, using the insights gained from this study.

References

[1] Cao, Meng & Yuan, Jinliang & Yu, Hualei & Zhang, Baoming & Wang, Chongjun. (2023). Self‐supervised short text classification with heterogeneous graph neural networks. Expert Systems. 10.1111/exsy.13249.

[2] Hussein, Hussein & Anwar, Said. (2021). Imbalanced Data Classification Using Support Vector Machine Based on Simulated Annealing for Enhancing Penalty Parameter. Periodicals of Engineering and Natural Sciences (PEN). 9. 1030. 10.21533/pen.v9i2.2031.

[3]Kumar, Chhotelal & Kumar, Mukesh. (2022). Comparative Analysis of Emotional State Classification Using Different Machine Learning Techniques. 10.1201/9781003145004-15.

[4] Joseph, Briskilal & Subalalitha, C.. (2021). Classification of Idioms and Literals Using Support Vector Machine and Naïve Bayes Classifier. 10.1007/978-981-16-5078-9_42.

[5] Mendoza, Marcelo & Zamora, Juan. (2009). Identifying the Intent of a User Query Using Support Vector Machines. 131-142. 10.1007/978-3-642-03784-9_13.

[6] X. Liu, Y. Chen, Y. Liu, and J. Tang, "Attention-based LSTM for target-dependent sentiment classification," in Proceedings of the Annual Meeting of the Association for Computational Linguistics, 2016.

[7] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2018.

[8] Liu, Hao & Peng, Huaming. (2021). Using BERT Model for Intent Classification in Human-Computer Dialogue Systems to Reduce Data Volume Requirement. 10.1007/978-3-030-80285-1_59.

[9] Liu, Yuanxia & Hao, Tianyong & Liu, Hai & Mu, Yuanyuan & Weng, Heng & Wang, Fu Lee. (2023). OdeBERT: One-stage Deep-supervised Early-exiting BERT for Fast Inference in User Intent Classification. ACM Transactions on Asian and Low-Resource Language Information Processing. 10.1145/3587464.

[10] Comi, Daniele & Christofidellis, Dimitrios & Piazza, Pier & Manica, Matteo. (2022). Z-BERT-A: a zero-shot Pipeline for Unknown Intent detection. 10.48550/arXiv.2208.07084.