Semi-Supervised Chinese Intent Classification using GAN-BERT

Aman Sharma
University of Waterloo
a74sharm@uwaterloo.ca

Marcus Huang
University of Waterloo
m43huang@uwaterloo.ca

Rithic Kumar Nandakumar
University of Waterloo
rknandak@uwaterloo.ca

Shriya Vaagdevi Chikati
University of Waterloo
svchikat@uwaterloo.ca

April 08, 2024

Abstract

This project aims to explore the capabilities of GAN-BERT model to address the challenges of semi-supervised intent classification, specifically within the Chinese language. The task of intent classification heavily relies on large, labelled datasets, which require intensive resources like time and effort to gather. To overcome this challenge, we use GAN, that generates a synthetic labelled dataset, alongside BERT, that provides with a feature extractor. This model uses both labelled and unlabelled data to aid with intent classification in natural language processing. We experimented with different ratios of labelled and unlabelled data to evaluate the model under varying degrees of supervision. The results highlight the model's robustness and its potential to reduce the reliance on huge, labelled datasets.

Keywords— Semi-Supervised Learning, Intent Classification, GAN, BERT

1 Introduction

Over the past few years, there has been a surge in the use of chatbots and virtual assistants, which has been impacted by the recent advancements in Artificial Intelligence. There has also been an increase in the quality of how these chatbots and assistants understand complex user queries and respond to them appropriately. As it is crucial for businesses to improve their customer engagement, these conversational agents have been pivotal to automating communication and it is vital for them to do so effectively.

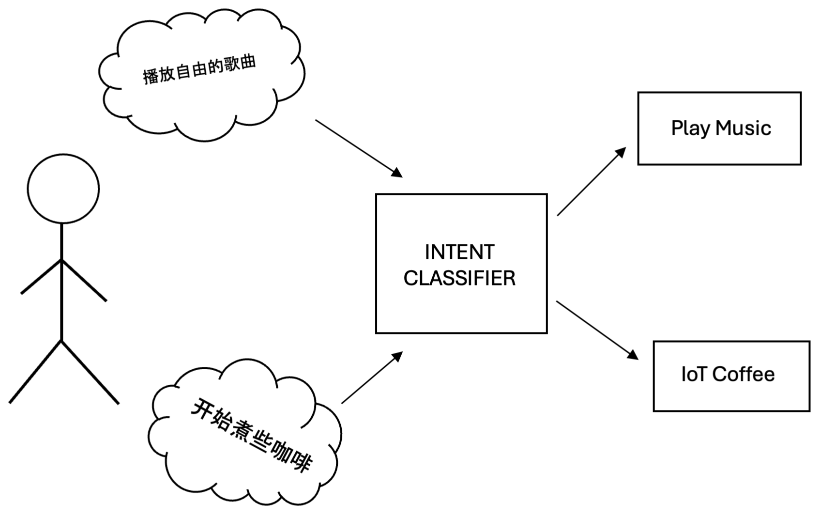

Figure 1: Chinese intent classification

The ability of these agents to understand user queries and respond accordingly depends on user intent classification. Intent classification, in natural language processing, takes in user queries as input and classifies it into one of the several classes to provide meaningful and contextually appropriate responses. This task relies heavily on huge, labelled datasets and gathering such dataset is often costly and time-consuming, which becomes even more problematic when there is a scarcity of such data.

This is where semi-supervised learning comes into play. Semi-supervised learning utilizes both labelled and unlabelled data as input, which eliminates the need for extensive labelled datasets. A model that has proven to be revolutionary for this specific task is GAN-BERT, which is an integration of the pre-trained Bidirectional Encoder Representations from Transformers (BERT), that requires a lot of labelled data to provide better results, and Generative Adversarial Network (GAN), that generates synthetic data thereby diminishing the need for gathering extensive labelled data.

We use the GAN-BERT model for intent classification specifically for the Chinese language and Figure 1 provides an example. For our project, we acquired the intent classification dataset, from Hugging Face, called MASSIVE. Our experiments include comparing different models trained on various ratios of labelled and unlabelled data to observe how the model performs under varying levels of supervision. The promising results from our experiments on the GAN-BERT model demonstrate its ability to perform relatively well even with being trained on enough unlabelled data coupled with labelled data. This paves way for implementing semi-supervised learning techniques in underrepresented languages that lack enough labelled data.

2 Related Works

Intent classification involves understanding the intent behind user utterances [6]. It plays a pivotal role in developing task-oriented dialog systems and so, with the surge in dialog systems, there has been a rise in the number of works in this area. The study by Chen et.al [1] introduced a model that combines intent classification and slot filling, based on BERT. They evaluated the model on two datasets and compared the results with BERT-based model and across various number of epochs. Their model performed relatively well.

Our work is related to implementing one of the semi-supervised learning strategies for intent classification. The field of semi-supervised learning has gained considerable attention because of its effectiveness in utilizing both labelled and unlabelled data to improve generalization capabilities using such data [2]. There has been extensive research in this field as summarized by the review by Ouali et. al [9], which outlines the dominant semi-supervised approaches used in deep learning.

We use the GAN-BERT model, which was proposed by Croce et. al [2], as a solution to the lack of vast amounts of labelled datasets and to use unlabelled datasets in text classification. Their evaluations across various datasets and multiple different tasks, demonstrate promising results even when the model is trained on few labelled examples. These results show that that GAN-BERT performs well, even with a small percentage of annotated data, compared to the original BERT model [2].

This study by Hasan et.al [7] introduced a semi-supervised GAN architecture model, built on top of a BERT model called GAN-BnBERT, for intent classification in the Bengali language. They developed a Bengali intent dataset, named BNIntent, by translating utterances from the CLINC150 dataset into Bengali, and evaluated their model on this dataset. Their findings included comparisons with previous results of intent classification in English, and the proposed model appears to provide comparable results. [7]"

3 Dataset

The dataset used for this project was obtained from MASSIVE [4]. MASSIVE is short for Multilingual Amazon SLU (Spoken Language Understanding) Resource Package for Slot-filling, Intent Classification, and Virtual Assistant Evaluation. It contains over 1 million parallel utterances across 52 languages, covering 18 domains, with 60 intents, and 55 slot types. We sourced the intent classification 'zh-CN' dataset, which stands for 'Chinese - China', from Hugging Face under the identifier 'AmazonScience/massive'. This dataset contains 16.5K rows.

4 Methodology

A Generative Adversarial Network (GAN) [5] architecture is combined with a pre-trained BERT (Bidirectional Encoder Representations from Transformers) [3] base Chinese model for intent classification with a semi-supervised approach. Semi-supervised learning combines both labelled and unlabelled data for training. This approach is particularly useful when there is a scarcity of labelled training data as seen in GAN-BERT [2].

4.1 Architecture

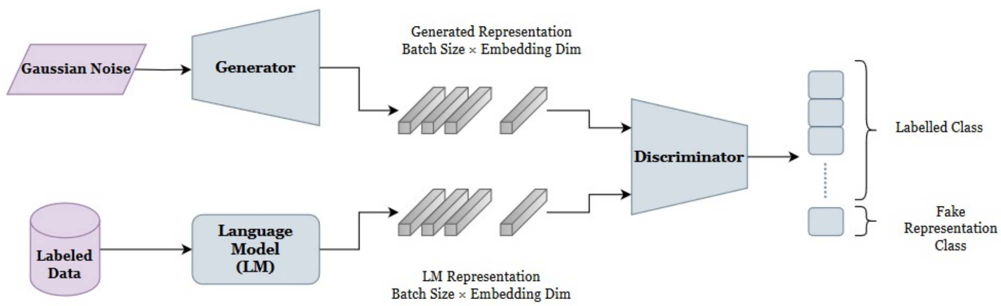

Figure 2: Architecture

The architecture can be divided into three components namely:

Generator block: This block generates synthetic/fake text samples from a Gaussian noise input. The input to the generator is a 100-dimensional noise vector. The noise vector is then passed through a hidden layer with LeakyReLU activation and dropout regularization. The output from this block is of the same dimensions as the hidden size of the transformer.

Discriminator block: This block takes the output representations from generator and BERT model and passes them through a hidden layer with LeakyReLU activation and dropout regularization to classify the input text as real or fake. In the case of real text, the corresponding intent is also predicted. The output size from this block is the number of intents + 1, where the extra dimension is for the classification of the sample being fake or real.

Language Model: A pre-trained BERT base-Chinese model is chosen as the language model for this task. This model is trained on a large corpus of Chinese text data making it suitable for our task. The architecture of this model comprises of 12 transformer encoder layers, with 768 hidden units and 12 attention heads. The maximum sequence length is set as 64. The model takes both labelled and unlabelled real data and the output representations are passed to the discriminator block for classification.

4.2 Training

The GAN is trained in an adversarial manner, where the generator tries to generate text similar to real data and the discriminator tries to distinguish between real and fake text as well as classify the intents in the case of real texts.

The generator's loss is the addition of two components namely the discriminator output for fake samples and the feature regularization loss. The discriminator output loss aims to push the generator to produce samples that maximize the probability of being classified as real by the discriminator. The feature regularization pushes the generator to produce samples whose features are similar to those of features obtained from real text.

The discriminator's loss is the sum of two losses namely the supervised and unsupervised loss. Supervised loss measures the loss in assigning the wrong intent to labelled data whereas unsupervised loss measures the error in wrongly classifying the fake texts as real and vice versa.

Both the generator and discriminator use AdamW (Adam with a weight decay) [8] with a learning rate of 5e-5. The batch size is set as 64.

5 Results

Since the MASSIVE dataset does not contain unlabelled data, we mask the labels of a proportion of the training set as unlabelled data and evaluate the performance based on different variants.

Table 1: Training Configurations

Variant	**Percentage masked (unlabelled)**	**Number of epochs**
1	0.9	20
2	0.8	18
3	0.6	16
4	0.4	14
5	0.2	12
6	0.1	10

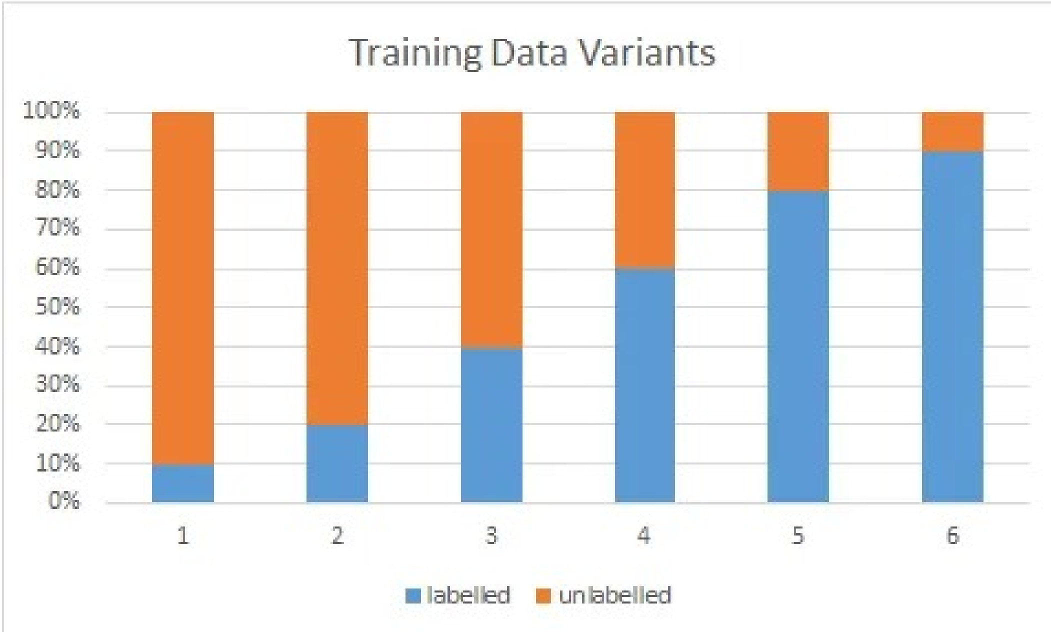

Figure 3: Training data variants

The intent classification accuracy of these variants on the test set is observed in table 2.

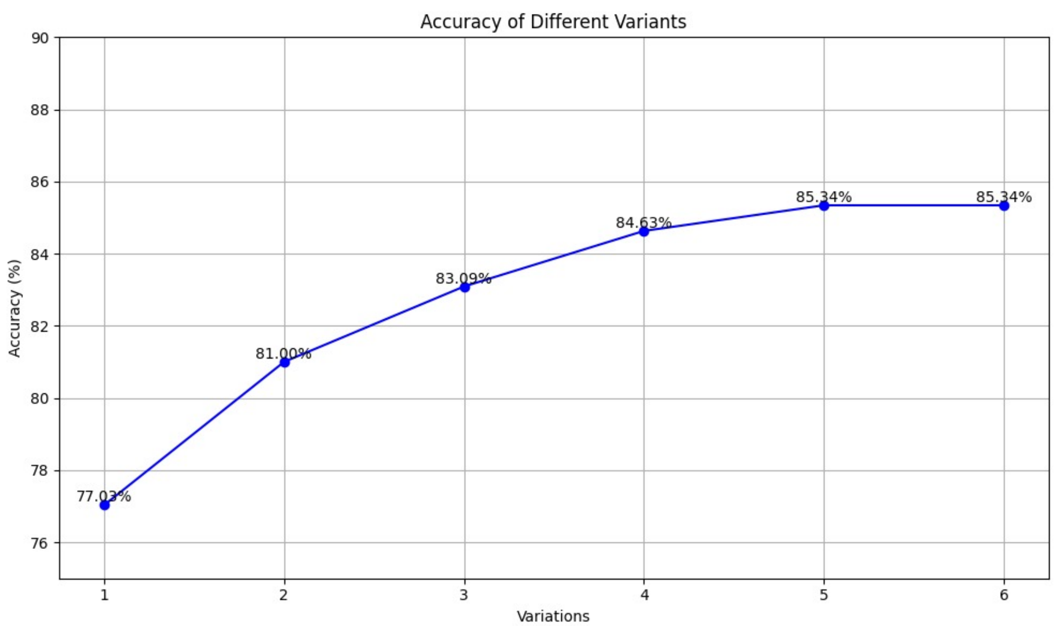

Figure 4: Intent Accuracy of variants

Table 2: Model Accuracy

Variant	Accuracy
1	77.03
2	81.00
3	83.09
4	84.63
5	85.34
6	85.34
Full dataset	85.4

It is observed that even with only 10% labelled data and 90% unlabelled data (Variant 1), 77.03 % intent classification accuracy is attained, compared to 85.4% accuracy when the full dataset is labelled. So the GAN-BERT model can classify intents using limited labelled training data with reasonable accuracy.

6 Conclusion

This study represents a pivotal advancement in the realm of Intent classification especially for those languages where labelled data is very limited.

Effective representation learning is made possible by the GAN-BERT architecture, which trains a discriminator to discern between actual and produced samples and a generator to generate realistic data samples. This adversarial training paradigm enables the model to capture latent features from both labelled and unlabelled data, enhancing its ability to generalize to unseen examples.

In many real-world scenarios, labelled data for specific tasks may be limited or expensive to acquire. By leveraging unlabelled data in a semi-supervised setting, our approach mitigates the effects of data sparsity, enabling the model to achieve competitive performance with less labelled data. This is particularly advantageous for low-resource languages like Chinese, where annotated datasets may be scarce.

The effectiveness of our suggested approach holds great promise for a number of applications, such as virtual assistants, chatbots, and customer support platforms, where precise intent classification is essential for comprehending user inquiries and offering pertinent answers. By leveraging semi-supervised learning techniques, organizations can effectively leverage existing data resources to improve the performance of their natural language understanding systems.

While our study demonstrates the effectiveness of semi-supervised intent classification with the GAN-BERT model, there are several avenues for future research. Investigating different strategies for incorporating unlabelled data, exploring domain adaptation techniques, and evaluating the robustness of the model to noisy or adversarial inputs are potential directions for further study.

In conclusion, our work underscores the importance of leveraging semi-supervised learning approaches for intent classification in Chinese, showcasing the utility of the GAN-BERT model in this context. By harnessing both labelled and unlabelled data, our approach offers a practical solution to the challenges of data scarcity and improves the performance of natural language understanding systems in real-world applications.

References

[1] Chen, Q., Zhuo, Z., and Wang, W. Bert for joint intent classification and slot filling, 2019.

[2] Croce, D., Castellucci, G., and Basili, R. GAN-BERT: Generative adversarial learning for robust text classification with a bunch of labeled examples. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (Online, July 2020), D. Jurafsky, J. Chai, N. Schluter, and J. Tetreault, Eds., Association for Computational Linguistics, pp. 2114–2119.

[3] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.

[4] FitzGerald, J., Hench, C., Peris, C., Mackie, S., Rottmann, K., Sanchez, A., Nash, A., Urbach, L., Kakarala, V., Singh, R., Ranganath, S., Crist, L., Britan, M., Leeuwis, W., Tur, G., and Natarajan, P. Massive: A 1m-example multilingual natural language understanding dataset with 51 typologically-diverse languages, 2022.

[5] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Generative adversarial networks, 2014.

[6] Gupta, A., Hewitt, J., and Kirchhoff, K. Simple, fast, accurate intent classification and slot labeling for goal-oriented dialogue systems, 2019.

[7] Hasan, M., Basher, M. J. I., and Shawon, M. T. R. Bengali intent classification with generative adversarial bert, 2023.

[8] Loshchilov, I., and Hutter, F. Decoupled weight decay regularization, 2019.

[9] Ouali, Y., Hudelot, C., and Tami, M. An overview of deep semi-supervised learning. *ArXiv abs/2006.05278* (2020).