Alexa utterances and intent mapping using BERT

Carla Prieto Chavez
UC Berkeley
MIDS 266 Final Project Presentation

Voice-activated virtual assistants, like Amazon Alexa, have revolutionized human-computer interaction by responding in real-time to user instructions, known as "utterances."

- Utterances mapped to specific "intents."
- In this project I explore a path in which Alexa processes and maps utterances to intents using different NLP techniques.
- According to a 2018 paper by Amazon Science, Alexa uses LSTM and personalized attention.
- For this project I used the MASSIVE dataset from AmazonScience.

Objective

Investigate the use of BERT models for intent recognition.
Leverage prior experience with BERT and engineering background with Alexa.
Experiments conducted on Google Colab due to resource constraints.

Why BERT?

- Leading model in NLP due to bidirectional sentence processing.
- Demonstrated significant performance improvements in various tasks (Devlin et al., 2019).
- Used in text classification tasks widely
- Pretrained version available from HuggingFace

Alexa's Intent Mapping

Process Overview

1. Automatic Speech Recognition (ASR): Converts spoken utterance into text.
2. Natural Language Understanding (NLU): Analyzes text to determine intent and extract relevant entities.
3. Intent Mapping: Matches analyzed text to a predefined intent using ML models.

Methodology

Exploratory Data Analysis (EDA)

- Dataset Length: Training and test sets.
- Utterance Length: Average length in characters and words.
- Distribution of Utterances per Intent: Examined to ensure equal distribution in training and test sets.

Approach

1. **BERT with Pooler Output**: Standard for using BERT.
2. **BERT with Averaged Output**: Uses average of the last hidden state outputs.
3. **BERT + CNN**: Attaches a CNN to process the output from BERT.

Model Results

First Run

Model	Runtime	Loss	Accuracy	Validation Loss	Validation Accuracy
BERT w/ Pooler Output	25m	1.6619	0.5906	3.6678	0.4469
BERT w/ Averaged Output	24.76m	1.0586	0.7440	3.4003	0.4839
BERT + CNN	47.35m	1.2624	0.6950	3.2038	0.4785

Baseline(Success) Defined as Validation Accuracy > 0.5

Model Results

Final Run

Model	Epochs	Runtime	Loss	Accuracy	Validation Loss	Validation Accuracy
BERT w/ Pooler Output	2	19.41m	0.5706	0.8688	3.4729	0.4778
BERT w/ Averaged Output	3	73.44m	0.3624	0.9114	4.7088	0.4667
BERT + CNN	2	4.25h	0.8156	0.8016	0.7981	0.7946

Takeaways

- Best Model: BERT + CNN, achieving validation accuracy above 0.50.
- Increasing the number of epochs does not necessarily lead to better results.
- Importance of validation metrics in model tuning and selection.
- Resource constraints and complexity of task influence model performance.

References

1. Kim, Y.-B., Kim, D., Han, K., and Lee, Y. (2018). Efficient Large-Scale Neural Domain Classification with Personalized Attention. Retrieved from https://pages.cs.wisc.edu/~ybkim/paper/acl2018.pdf
2. Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of Deep Bidirec- tional Transformers for Language Understanding. Retrieved from https://arxiv.org/abs/1810.04805
3. Amazon Science (2023). AmazonScience/massive Dataset. Retrieved from https://huggingface.co/ datasets/AmazonScience/massive
4. Github repo with calculations and models: https://github.com/Carla08/alexa_utt_intention_bert