**Abstract**

Nowadays, Machine Learning has been widely used in customer service area. It significantly reduced total cost of ownership via reducing human interactions. In this scenario, the first and the most useful step is to predict the intent of a customer contact. The problem we are trying to solve in this project is to detect intent from an utterance in different language, including English, French, German, Spain, and Chinese. The input of our model in training phase is in one language - English, and that in inference phase is sentences from other 4 languages. The output is a ranking of a list of pre-defined intents, such as greeting, negative feedback, taking to agent or asking for refund status, etc.

**Dataset**

- **Training data**: We are using customer support data collected from a real customer service channel with sensitive information removed. This data set has about 3.5k sentences and the corresponding intent labeled by human. The data is split into 2.7k training set and 800 test set. There are 11 different intents in this dataset.

- **Data Sample and Label Distribution**:

  - **Label Distribution For Training Set**:

    - **Label**: greeting, negfdbk, negfdbk, negfdbk
    - **Data Sample and Label Distribution**:
      - negfdbk: facilitas el contacto

- **Data Processing**

  - **Training Data Sample**:
    - training data: [Image]
    - testing data: [Image]

- **Rapid Model Creation**

  - To speed up our experiment, we have built a UI based tool named NLP Model Factory which allows the users to choose base model (BERT in our example) for transfer learning / fine tune, specify hyper parameters, submit the job to GPU boxes, evaluate the result, generate reports and track jobs.

**Methodology**

We've tried two approaches on this multi-class intent prediction problem:

1. **Approach 1**: We use BERT Base Model as embedding layer and added FC layers to do a fine tune. The model we created can only understand English input data. Later on in prediction phase, we use Google machine translation to translate the sentences of different languages into English and then use our model to predict the intent.

   - **Training**:
   - **Inference**:

2. **Approach 2**: We use BERT Multilingual model which was trained on multi-language corpus. In training phase we apply English training data to create a model that can understand our problem in multiple language. Finally we pass in test data in other language and let the model predict the intent directly. This is a zero shot learning.

   - **Training**:
   - **Inference**:

**Model Performance**

We evaluate the model performance by primarily looking at overall accuracy, and then precision and recall of each intent.

<table>
<thead>
<tr>
<th>Language</th>
<th>Approach</th>
<th>Acc</th>
<th>Refund</th>
<th>Agent</th>
<th>Hello</th>
<th>Negfdbk</th>
</tr>
</thead>
<tbody>
<tr>
<td>EN1</td>
<td>94.7</td>
<td>96.8</td>
<td>100.0</td>
<td>90.3</td>
<td>93.5</td>
<td>87.5</td>
</tr>
<tr>
<td>EN2</td>
<td>92.8</td>
<td>96.5</td>
<td>99.0</td>
<td>89.3</td>
<td>94.3</td>
<td>75.0</td>
</tr>
<tr>
<td>DE1</td>
<td>93.8</td>
<td>97.4</td>
<td>99.0</td>
<td>92.7</td>
<td>85.7</td>
<td>60.0</td>
</tr>
<tr>
<td>FR1</td>
<td>94.3</td>
<td>98.0</td>
<td>99.0</td>
<td>91.5</td>
<td>95.2</td>
<td>85.7</td>
</tr>
<tr>
<td>DE2</td>
<td>80.1</td>
<td>92.7</td>
<td>88.1</td>
<td>93.4</td>
<td>68.5</td>
<td>100.0</td>
</tr>
<tr>
<td>DE2</td>
<td>80.1</td>
<td>92.7</td>
<td>88.1</td>
<td>93.4</td>
<td>68.5</td>
<td>100.0</td>
</tr>
</tbody>
</table>

**Conclusion**

The result shows currently the Approach 1 has achieved better overall performance than Approach 2. We have done some analysis and here is our findings:

1. Single language model performs better in general than multi-language model for testset in a particular language that this model is training with.
2. We also noticed that the test data is from machine translation which is quite different from real human conversation in other language, however, it’s friendly for machine translation engine to translate it back to English. This gives advantage in the evaluation for Approach 1 over Approach 2.
3. Multi-language model has shown solid result from zero-shot learning. This has significant advantage regarding to cost and time to market.

**Future Work**

As next steps, we have following proposals:

1. Validate the result with more data and more accurate data. one thing we can do is data augmentation, another is to collect real chat messages in other languages and use them as test data.
2. Architecture search. We can extract more features from BERT embedding layers, and try more architecture for FC part.
3. Pre-training on BERT model. there is some data that currently pre-trained model provided by google can not handle because they may not present in the corpus, such as some company or industry specific terminology and their translations.

**Acknowledgements**

Special thanks to Charles(Yancheng) Liu who provided technical support of GPU based training environment.