* FastText-Based Intent Detection for Inflected Languages:

<https://www.mdpi.com/2078-2489/10/5/161/pdf-vor>

Rating: 6/10

Advantages:

Uses pre-trained word embeddings (FastText) which break words into sub-word units and thus can account for out of vocabulary words. (May use an FC or Convolutional NN for classification).

Pre-trained on a larger text corpus and hence alleviate the need for large datasets, as a smaller dataset can be used for fine-tuning.

Disadvantages:

Does not take into account the full context of the sentence as it makes use of only the average of word embeddings to evaluate the total sentence embedding, which cannot give importance to the sequence of words/ context of the sentence. Therefore may produce inefficient sentence embeddings and incorrect intent classification.

* Intent Detection with WikiHow:

<https://arxiv.org/pdf/2009.05781.pdf>

<https://github.com/zharry29/wikihow-intent>

Rating: 5/10

Advantages:

Suggests using WikiHow as a source of text corpus for labelled intent detection data, for pre-training models. This helps ease the problem of unavailability of large amounts of data due to wide ranging domains and languages of WikiHow articles.

The huge corpus extracted from WikiHow makes open-domain intent detection and zero-shot intent detection (with not optimal but sufficient accuracy) a possibility.

Disadvantages:

This paper does not introduce a novel model architecture, rather it highlights a way of obtaining data and the advantages of pre-training of a wide dataset.

* Towards Open Intent Discovery for Conversational Text:

<https://arxiv.org/pdf/1904.08524.pdf>

Rating: 7/10

Advantages:

TOP-ID, can discover both previously seen as well as unseen (during training) user intents in diverse real-world scenarios (not restricted to a pre-defined set of intent categories).

Can identify multiple user intents per utterance.

Presents a large, intent-annotated dataset of 25K text instances from real-world task domains (extracted from Stack Exchange), without any restriction on the number or types of intents possible.

Disadvantages:

Focuses on identifying and extracting user intentions from text utterances explicitly containing them in their content. It does not infer or deduce the intent if it is implicitly stated in the text.

Two separate models need to be trained for the two-step task (detecting whether intent exists (binary classification), and identifying the action and object (sequence tagging) that comprise the intent).

* Attention-Based Recurrent Neural Network Models for Joint Intent Detection and Slot Filling:

<https://arxiv.org/pdf/1609.01454v1.pdf>

Rating: 8/10

Advantages:

Facilitates and improves performance of Slot filling and Intent Detection by sharing some data between the tasks by sharing a common encoder.

Disadvantages:

Shares only the encoder representations of the Utterances between the two tasks which does not take full advantage of the dependency of slot filling on intent and vice versa.

* BERT for Joint Intent Classification and Slot Filling:

<https://arxiv.org/pdf/1902.10909v1.pdf>

Rating: 7/10

Advantages:

Uses pre-trained word embeddings (BERT) which break words into sub-word units and thus can account for out of vocabulary words.

BERT embeddings take into account the context that the word has been used in in the specific utterance to a greater extent.

Pre-trained on a larger text corpus and hence alleviate the need for large datasets, as a smaller dataset can be used for fine-tuning.

Disadvantages:

Does not take into account the full context of the sentence as it makes use of only the cls token embeddings to evaluate the total sentence embedding (cls token is a good way of representing sentence embeddings however further processing with LSTM models produce better sentence embeddings). Therefore may produce inefficient sentence embeddings and incorrect intent classification.

* Deep Bi-Directional LSTM Network for Query Intent Detection:

<https://www.sciencedirect.com/science/article/pii/S1877050918320374>

Rating: 8/10

Advantages:

Uses word embeddings (may use pre-trained), and a deep Bi-LSTM network which can capture the contextual and semantic properties of a sentence leading to better intent classification/detection.

* INTENT DISCOVERY THROUGH UNSUPERVISED SEMANTIC TEXT CLUSTERING:

<https://www.isca-speech.org/archive/Interspeech_2018/pdfs/2436.pdf>

Rating: 7/10

Advantages:

Unsupervised, therefore alleviates the need for large datasets for training/testing.

Disadvantages:

Less efficient that supervised methods as it may not account for many properties of individual utterances as it works entirely on the basis of clustering.

* Unsupervised Classification of Student Dialogue Acts With Query-Likelihood Clustering:

<https://research.csc.ncsu.edu/learndialogue/pdf/LearnDialogue-EzenCan-EDM-2013.pdf>

Rating: 7/10

Advantages:

Unsupervised, therefore alleviates the need for large datasets for training/testing.

Disadvantages:

Less efficient that supervised methods as it may not account for many properties of individual utterances as it works entirely on the basis of clustering.

* A Stack-Propagation Framework with Token-Level Intent Detection for Spoken Language Understanding:

Rating: 9/10

Advantages:

Facilitates and improves performance of Slot filling and Intent Detection by sharing data between the tasks by sharing a common encoder to a greater extent. The results of intent detection directly impact the process of slot filling using stack propagation.

Performs token level intent detection, can provide features at each token to slot filling in Stack-Propagation framework, which can ease the error propagation and retain more useful information for slot filling. Compared with sentence level intent detection, if the intent of the whole sentence is predicted wrongly, the wrong intent would possibly apply a negative impact on all slots. However, in token-level intent detection, if some tokens in the utterance predicted wrongly, other correct token-level intent information will still be useful for the corresponding slot filling

Since each token can grasp the whole utterance contextual information by using the self-attentive encoder, we can consider predictions at each token in an utterance as individual prediction to the intent of this utterance, this approach will reduce the predicted variance and improve the performance of intent detection.

Can be used along with BERT embeddings model (replace encoder with BERT) to produce even better results and further alleviate the need for large data corpus.

* Intention Detection Based on Siamese Neural Network With Triplet Loss:

<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9082602>

Rating: 9/10

Advantages:

Adopts a two-stage process for intent classification:

1. Feature embedding learning using BERT/RMCNN combined with triplet loss function (semantically/contextually similar utterances end up with similar embeddings).
2. Intent detection (the features from 1 are fed into a softmax layer which classifies the intent).

Generalizes better and can distinguish between the same utterance with multiple possible tags depending on context using the feature embeddings generated using triplet loss.

Disadvantages:

Have to put extra effort for triplet sampling.