Dialogue Act Classification Exploiting Semantic and Contextual Information

Dimitris Samouil, Apostolos Gnardellis, Petros Hadjikkos
Department of Computer Science
UCL

Introduction

The notion of Dialogue Act (DA) plays a crucial and leading role in the natural language processing field. The concept of DA initially developed by Austin (1962). In recent years, several artificial neural networks (ANNs) methods present very promising results for short-text classification tasks.

The aim of this work has two sub-goals:

- Firstly, an investigation of the procedure of DA classification and a detailed literature review related to methods that has been applied to it, are presented.
- Secondly, an innovative sequential utterances classification model, combined with a reinforced sentence representation model, is introduced.

Methods

LSTM Utterance Representation

For each utterance which contains k words, we represent the sentence as a sequence of k word vectors $x_1, x_2, \ldots, x_k$ with dimension m. These word vectors are given as inputs to the LTSM structure, as shown in *Figure 1*, to produce a l-dimensional vector that represents the utterance. For a given t^{th} word in the utterance, an LSTM cell takes as inputs x_t, h_{t-1}, c_{t-1} and computes h_t, c_t as follows:

$$I_t = \sigma(W_i x_t + U_i h_{t-1} + b_i)$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o)$$

$$u_t = \tanh(W_u x_t + U_u h_{t-1} + b_u)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot c_t$$

$$h_t = o_t \odot \tanh(c_t)$$

Sequential Utterance Classification

The act of each utterance is predicted using the RNN architecture shown in *Figure 2*. Let s_t be the l-dimensional vector that represents the t^{th} utterance of a dialogue computed by the LSTM model. The sequence s_{t-1}, s_{t-2}, s_t is fed to the RNN and the output is the dialogue act of the utterance st computed as follows :

$$h_t = \tanh(W_s x_t + W_h h_{t-1} + b)$$

$$act_t = \frac{e^{Vhs+b}}{\sum_s e^{Vhs+b}} = g(h_t + b)$$

Word Semantic Representation

In this task, we attempted to reinforce the typical word embeddings by adding extra information in each word vector. This model provides information about the correlation between each word and each DA act by computing their semantic similarity, which then can be incorporated in the pre-trained word vectors.

$$p(t_i, w) = \frac{1}{n}\sum_{i=1}^{n} p(t_i|w)p(w)$$

$$p(w|t_i) = \frac{p(w,t)}{\sum w' \in Vocabulary\, p(w',t)}$$

$$s(w, t_i) = \sum_{j=1}^{N} \frac{p(t_i|k_j)p(k_j)}{p(t_i)}\, Dc(k_j w)$$

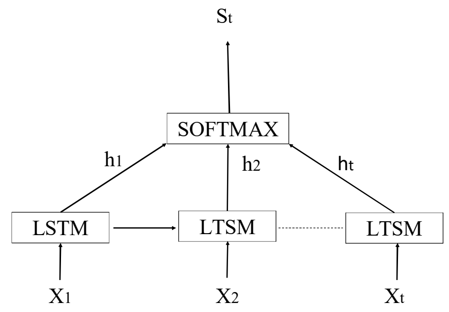

Figure 1: LSTM model for Utterance Representation

DA act	Example
Statement-non-opinion	*Me, I'm in the legal department*
Acknowledge (Backchannel)	*Uh-huh*
Statement-opinion	*I think it's great*
Agree/Accept	*That's exactly it.*
Abandoned or Turn –Exit	*Do you,-*
Non-verbal	*[Laugher]*
Appreciation	*I can imagine.*
Yes-No-Question	*Do you have to have any special training*

Dataset

The model was evaluated on Switchboard (SwDA) dataset which is annotated with 42 DA acts (Jurafsky et al., 1998) and is split into the training set (1,155 dialogues of 199,050 utterances) and the test set (19 dialogues of 3,927 utterances). The included dialogues concern spontaneous human-to-human telephone conversations.

In the table, eight representative utterances of the most repeated DAs acts are shown.

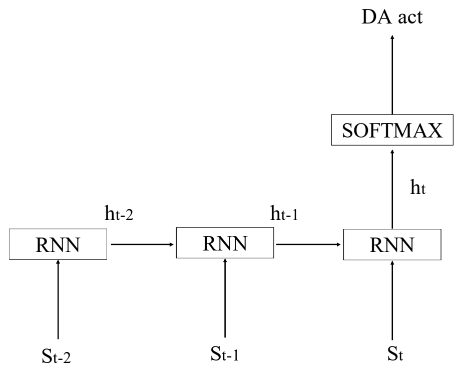

Figure 2: Sequential utterance classification

Model	Accuracy (%)
Majority classification baseline	*31.6*
HMM (Stolcke et al., 2000)	*71.0*
CNN (Lee and Dernoncourt, 2016)	*73.1*
DRLM – conditional training (Ji et al., 2016)	*77.0*
LTSM (Khanpour et al., 2016)	*80.1*
Semantic LSTM (Papalampidi et al., 2017)	*75.6*
Our baseline	***70.7***
Our baseline with reinforced embeddings	***73.8***
Proposed	***77.5***

Results

- The proposed model, namely the baseline with the reinforced (semantic) embeddings and the discourse model, achieved an accuracy of 77.5%, outperforming the majority of the corresponding models.
- Our original baseline model succeeded accuracy equal to 70.7%, while the baseline fed by the semantic embeddings led to an accuracy of 73.8%, meaning that the proposed model achieved higher accuracy by 3.8% and 6.8%, respectively.

Conclusion

In this work, we demonstrated the effectiveness of combination of the baseline model with a sequential utterances classification model, incorporated by reinforced (semantic) word embeddings. The results of the method showed an increase in the overall accuracy of the dialogue act classification, making it superior to the majority of the approaches from the literature. Regarding future work, it would be useful to examine the combination of higher number of features being exported from detailed discourse analysis. Additionally, it could be worth to determine the portability of the semantic model approach to various corpora.