Available online at www.sciencedirect.com

ScienceDirect

Computer Speech & Language 47 (2018) 175–193

www.elsevier.com/locate/csl

On the effects of using word2vec representations in neural networks for dialogue act recognition ☆

Christophe Cerisara[a,*], Pavel Král[b,c], Ladislav Lenc[b,c]

[a] LORIA-UMR7503, Campus Scientifique, 54506 Vandoeuvre-les-Nancy, France
[b] Department of Computer Science & Engineering, Faculty of Applied Sciences, University of West Bohemia, Plzeň, Czechia
[c] NTIS – New Technologies for the Information Society, Faculty of Applied Sciences, University of West Bohemia, Plzeň, Czechia

Received 11 March 2016; received in revised form 17 March 2017; accepted 26 July 2017
Available online 29 July 2017

Abstract

Dialogue act recognition is an important component of a large number of natural language processing pipelines. Many research works have been carried out in this area, but relatively few investigate deep neural networks and word embeddings. This is surprising, given that both of these techniques have proven exceptionally good in most other language-related domains. We propose in this work a new deep neural network that explores recurrent models to capture word sequences within sentences, and further study the impact of pretrained word embeddings. We validate this model on three languages: English, French and Czech. The performance of the proposed approach is consistent across these languages and it is comparable to the state-of-the-art results in English. More importantly, we confirm that deep neural networks indeed outperform a Maximum Entropy classifier, which was expected. However, and this is more surprising, we also found that standard word2vec embeddings do not seem to bring valuable information for this task and the proposed model, whatever the size of the training corpus is. We thus further analyse the resulting embeddings and conclude that a possible explanation may be related to the mismatch between the type of lexical-semantic information captured by the word2vec embeddings, and the kind of relations between words that is the most useful for the dialogue act recognition task.

Keywords: Dialogue act; Deep learning; LSTM; Word embeddings; Word2vec

1. Introduction

1.1. Dialogue act recognition

Mutual understanding in interactive situations, either when several people are engaged in a dialogue or when they are interacting with a modern computer system in natural language, may not be achieved without considering both the semantic information in the speakers utterances and the pragmatic interaction level,

☆ This paper has been recommended for acceptance by Roger K. Moore.

* Corresponding author.
E-mail address: cerisara@loria.fr (C. Cerisara), pkral@kiv.zcu.cz (P. Král), llenc@kiv.zcu.cz (L. Lenc).

http://dx.doi.org/10.1016/j.csl.2017.07.009
0885-2308/

especially relative to dialogue acts (Sridhar et al., 2009). Dialogue Acts (DAs) represent the meaning of an utterance (or its part) in the context of a dialogue (Austin, 1962; Bunt, 1994), or, in other words, the function of an utterance in the dialogue. For example, the function of a question is to request some information, while an answer shall provide this information. Dialogue acts are thus commonly represented as phrase-level labels such as statements, yes-no questions, open questions, acknowledgements, and so on.

Automatic recognition of dialogue acts is a fundamental component of many human-machine interacting systems that support natural language inputs. For instance, dialogue acts are typically used as an input to the dialogue manager to help deciding on the next action of the system: giving information when the user is asking a question, but eventually keeping quiet when the user is just acknowledging, giving a comment, or even asking for delaying the interaction. In the latter case, a system reaction may be perceived as intrusive. Beyond human-machine interaction, this task is also important for applications that rely on the analysis of human-human interactions, either oral, e.g., in recordings of meetings (Zimmermann et al., 2006), or written, e.g., through the reply and mention-at structures in Twitter conversations (Ritter et al., 2010; Zarisheva and Scheffler, 2015; Vosoughi and Roy, 2016). It is also essential for a large range of other applications, for example talking head animation, machine translation (Fukada et al., 1998), automatic speech recognition or topic tracking (Garner et al., 1996). The knowledge of the user dialogue act is useful to render facial expressions of an avatar that are relevant to the current state of the discourse. In the machine translation domain, recognizing dialogue acts may bring relevant cues to choose between alternative translations, as the adequate syntactic structure may depend on the user intention. Automatic recognition of dialogue acts may also be used to improve the word recognition accuracy of automatic speech recognition systems, where a different language model is applied during recognition depending on the dialogue act (Stolcke et al., 1998).

To conclude, dialogue act recognition is an important building block of many understanding and interacting systems.

1.2. Motivation and objectives

Researches on dialogue act recognition have been carried out for a long time, as detailed in Section 2. The majority of these works exploit supervised learning with lexical, syntactic, prosodic and/or dialogue history features (Fišel, 2007). However, few approaches consider semantic features, while they may bring additional information and prove useful to improve the accuracy of the dialogue act recognition system. For instance, a frequent cause of recognition errors are "unknown" words in the testing corpus that never occur in the training sentences. Replacing specific named entities in the text (such as town names, dates...) by their category has been proposed in the literature as a remedy to this issue (Sanchis and Castro, 2002). We investigate a more general solution that exploits lexical similarity between word vectors. These word vectors may be computed in various ways, but they typically include mostly lexical semantic information about the word itself as well as some syntactic information, e.g., related to the relative position or degree of proximity of pairs of words within a sentence. This additional information may be used to improve dialogue act recognition, in particular when the training and test conditions differ, or when the size of the training corpus is relatively small.

In this work, we propose a new Deep Neural Network (DNN) based on Long Short-Term Memory (LSTM) for the task of dialogue act recognition, and we compare its performance to a standard Maximum Entropy model. Our first objective is to leverage the modelling capacity of such a DNN in order to achieve dialogue act recognition with only the raw observed word forms, i.e., without any additional expert-designed feature. This model is described in Section 3.2. The second objective is to further validate this model both on a standard English DA corpus, as well as on two other languages, without changing anything in the model, in order to assess the genericity and robustness of the approach. These experiments are summarized in Section 5. Finally, our third objective is to study the impact of word embeddings, which have been shown to provide extremely valuable information in numerous Natural Language Processing (NLP) tasks, but which have never been used so far [1] for dialogue act recognition. This study is summarized in Section 5.5.

[1] To the best of our knowledge at the time of submission.