Information Processing and Management 54 (2018) 657–673

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

Neural word and entity embeddings for ad hoc retrieval

Ebrahim Bagheri*[,a], Faezeh Ensan[b], Feras Al-Obeidat[c]

[a] Laboratory for Systems, Software and Semantics (LS3), Department of Electrical and Computer Engineering, Ryerson University, Canada
[b] Department of Computer Engineering, Ferdowsi University of Mashhad, Mashhad, Iran
[c] College of Technological Innovation, Zayed University, UAE

ARTICLE INFO

Keywords:
Neural embeddings
Ad hoc document retrieval
TREC
Knowledge graph

ABSTRACT

Learning low dimensional dense representations of the vocabularies of a corpus, known as *neural embeddings*, has gained much attention in the information retrieval community. While there have been several successful attempts at integrating embeddings within the ad hoc document retrieval task, yet, no systematic study has been reported that explores the various aspects of neural embeddings and how they impact retrieval performance. In this paper, we perform a methodical study on how neural embeddings influence the ad hoc document retrieval task. More specifically, we systematically explore the following research questions: *(i)* do methods solely based on neural embeddings perform competitively with state of the art retrieval methods with and without interpolation? *(ii)* are there any statistically significant difference between the performance of retrieval models when based on *word* embeddings compared to when knowledge graph *entity* embeddings are used? and *(iii)* is there significant difference between using locally trained neural embeddings compared to when globally trained neural embeddings are used? We examine these three research questions across both *hard* and *all* queries. Our study finds that *word embeddings* do not show competitive performance to any of the baselines. In contrast, *entity embeddings* show competitive performance to the baselines and when interpolated, outperform the best baselines for both hard and soft queries.

1. Introduction

The area of ad hoc document retrieval has received extensive treatment over the past several years whereby different retrieval models have been proposed to connect query and document spaces. Many of these works build on the foundations of language modeling techniques (Ponte & Croft, 1998) and offer variations that focus on certain aspects of the retrieval process such as impact of smoothing techniques (Zhai & Lafferty, 2001), integration of topic models (Wei & Croft, 2006), including external information in the retrieval process (Li, Luk, Ho, & Chung, 2007; Liu, Liu, Yu, & Meng, 2004), term dependency models (Huston & Croft, 2014), and deep neural networks (Guo, Fan, Ai, & Croft, 2016), just to name a few. More recently two additional directions have been recognized to have the potential to impact the document retrieval process, namely, the use of *knowledge graph* information as well as *neural embeddings*. Both of these techniques are focused on extending language models to move beyond a *hard match* between the query and document spaces, hence addressing issues such as the *vocabulary mismatch* problem.

Techniques based on knowledge graphs explore ways in which additional information related to the query or documents can be included in the retrieval process by systematically traversing through or summarizing the information content of the knowledge graph (Nikolaev, Kotov, & Zhiltsov, 2016; Xiong, Power, & Callan, 2017). As such, entity-centric retrieval models have been explored

* Corresponding author.
E-mail addresses: bagheri@ryerson.ca (E. Bagheri), ensan@um.ac.ir (F. Ensan).

https://doi.org/10.1016/j.ipm.2018.04.007
Received 4 November 2017; Received in revised form 10 March 2018; Accepted 17 April 2018

(Foley, O'Connor, & Allan, 2016; Hasibi, Balog, & Zhang, 2017) where entities represent knowledge graph concept mentions within the query or documents, which are often extracted using automated semantic annotators (Jovanovic et al., 2014). Given a set of entities, language models are either extended to support for entity information (Hasibi, Balog, & Bratsberg, 2016) or are interpolated with an additional language model built specifically for entities (Raviv, Kurland, & Carmel, 2016). A benefit of employing entities is they provide means for *soft matching* (Guo, Fan, Ai, & Croft, 2016) where semantic similarity measures (Zhu & Iglesias, 2017) can be used to calculate the distance of query–document pairs.

Neural embedding techniques provide a low dimensional yet dense vector representation of the terms while preserving the geometric relations between them; therefore, these methods provide similar benefits to those provided by the soft matching capability of knowledge graph entities (Ganguly, Roy, Mitra, & Jones, 2015). Various methods have been proposed that learn embeddings for documents (Le & Mikolov, 2014), words (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013), entities (Lin, Liu, Sun, Liu, & Zhu, 2015) as well as jointly for words and entities (Toutanova et al., 2015). Neural embeddings have been used in document retrieval for purposes such as query expansion (Diaz, Mitra, & Craswell, 2016), query classification (Zamani & Croft, 2017), ranking (Kuzi, Shtok, & Kurland, 2016) and text classification (Wang et al., 2016), to name a few. Given the growing role of neural embeddings of words and entities in the retrieval literature, it is important to understand their impact on the performance of ad hoc retrieval.

The challenge that are interested to address in this paper is that although earlier techniques in the literature have reported strong and systematic performance results on standard corpora, no methodical and comprehensive work is yet to comparatively report on the various aspects of neural embeddings in ad hoc retrieval. As such, it is not clear from a practical perspective *how* and to *what extent* neural embeddings can positively or negatively impact the ad hoc retrieval task. Therefore, this motivates us to systematically study the impact of neural embeddings on ad hoc retrieval from several cross-cutting aspects that have not been studied before, including (1) the impact of neural embeddings when learnt on words compared to when trained for knowledge graph entities, (2) the difference between learning neural embeddings at local and global scales, (3) the effect of neural embeddings on hard versus soft queries, and finally, the performance of neural embedding-based retrieval models compared to the state of the art.

More specifically, we study:

1. Whether there is any significant performance difference between the use of *word* or *entity* embeddings in the retrieval process or not. Several recent work have reported the impact of considering entities in ad hoc document retrieval. The observation has been that the consideration of entities and features learnt based on the context of entities within the knowledge graph can enhance the performance of keyword-based retrieval models, for instance, some authors have shown that when entities are present in the query, i.e., *entity-bearing queries*, it would be possible to retrieve effective yet non-overlapping documents to the keyword-based models (Ensan, Bagheri, Zouaq, & Kouznetsov, 2017; Liu & Fang, 2015). However, such distinction has not been studied in the embedding-based retrieval models. In other words, would the performance of a retrieval model that uses *word* embeddings be different from a model that uses *entity* embeddings and whether one of the two embedding types is more effective than the other.
2. If using *globally* trained embeddings results in noticeable difference compared to when the embeddings are *locally* trained for the specific retrieval task. Several authors have explored the impact of embeddings on the retrieval process with mixed results. For instance, Zuccon, Koopman, Bruza, and Azzopardi (2015) reported that globally trained embeddings improve the performance of a translation language model, while Diaz et al. (2016) reported that the use of local embeddings, trained based on relevant documents to the query collection, outperforms the performance of global embeddings. In constrast, Rekabsaz, Lupu, Hanbury, and Zamani (2017) argued that using global embeddings can lead to *topic drift*. Therefore, it is important to investigate the impact of global and local embeddings in the context of and in contrast to strong baselines.
3. Whether there is a significant difference on how embedding-based models impact *harder queries* (Carmel, Yom-Tov, Darlow, & Pelleg, 2006) compared to other queries. Some authors have discussed that some retrieval models such as LES (Liu & Fang, 2015) and SELM (Ensan & Bagheri, 2017) are more effective on harder queries, primarily because they identify relevant documents that suffer from the *vocabulary mismatch* problem. For this reason, we explore whether different embedding-based models impact the performance of hard queries differently from the others and as such would embeddings be most appropriately used within the context of such queries.
4. If the interpolation of an embedding-based model with a strong baseline shows any improvement over the state-of-the-art baselines. Most work in the literature have reported their findings of the performance of the embedding-based models when interpolated with a keyword-based retrieval model. For instance, Diaz et al. (2016) as well as Zamani and Croft (2017) interpolate their embedding models with a query language model based on Kullback–Leibler divergence between the query and document. On this basis, its valuable to explore whether there are cases when non-interpolated embedding-based models have competitive performance to the baselines and also would embedding-based models improve the performance of any baseline with which they interpolate with and would they always provide superior performance over a strong baseline.

In order to study these four aspects, we systematically performed experiments based on two large-scale TREC collections, namely ClueWeb'09B and ClueWeb'12B with their related topics (queries). In the experiments, we employ neural embedding representation of words and entities as a way to compute the distance between the query and document spaces. We use the Word Mover's Distance measure for the purpose of distance calculation between a query and a document based on their neural embedding representation. This distance is then used to rank document relevancy score given an input query. The produced rankings are evaluated based on gold standard human-provided relevance judgments already available in the TREC collection and then compared to several state-of-the-art baselines for comparative analysis.