Signal Processing 142 (2018) 403–411

Contents lists available at ScienceDirect

Signal Processing

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

Learning distributed sentence representations for story segmentation

Jia Yu[a,b], Lei Xie[a,*], Xiong Xiao[c], Eng Siong Chng[c]

[a] Shaanxi Provincial Key Laboratory of Speech and Image Information Processing, School of Computer Science, Northwestern Polytechnical University, Xi'an, China
[b] School of Computer and Information Engineering, Luoyang Institute of Science and Technology, Luoyang, China
[c] Temasek Laboratories@NTU, Nanyang Technological University, Singapore

ARTICLE INFO

Article history:
Received 31 March 2017
Revised 21 July 2017
Accepted 22 July 2017
Available online 24 July 2017

Keywords:
Distributed representation
Deep neural network
Topical sentence representation
Word vector
Sentence vector

ABSTRACT

Traditional sentence representations such as bag-of-words (BOW) and term frequency-inverse document frequency (tf-idf) face the problem of data sparsity and may not generalize well. Neural network based representations such as word/sentence vectors are usually trained in an unsupervised way and lack the topic information which is important for story segmentation. In this paper, we propose to learn sentence representation by using deep neural network (DNN) to directly predict the topic class of the input sentence. By using supervised training, the learned vector representation of sentences contains more topic information and is more suitable for the story segmentation task. The input of the DNN is BOW vector computed from a context window. Multiple time resolution BOW and bottleneck features (BNF) are also introduced to enhance the performance of story segmentation. As text data labeled with topic information is limited, we cluster stories into classes and use the class ID as the topic label of the stories for DNN training. We evaluated the proposed sentence representation with the TextTiling and normalized cuts (NCuts) based story segmentation methods on the topic detection and tracking (TDT2) task. Experimental results show that the proposed topical sentence representation outperforms both the BOW baseline and the recently proposed neural network based representations, i.e., word and sentence vectors.

1. Introduction

The exponential growth of human-centered media data makes the searching of needed information an increasingly difficult task. Content management techniques have been proposed to address the challenge, including topic detection and tracking [1,2], summarization [3], information extraction [4], and content indexing and retrieval [5]. Serving as an import preprocessing to the previously mentioned applications, story segmentation [6–8] is the task of automatically partitioning a stream of multimedia data into topically coherent segments. For example, human-centered social media data, e.g. chat logs from an instant messenger, usually contain multiple topics; segmenting these chat logs into topically coherent segments definitely benefits social media analysis. Story segmentation has been historically studied for diverse genres, such as broadcast news programs [9], meeting recordings [10] and lectures [11,12], over different types of media including audio [12–14], video [15] and text [6,8,10,16–18]. In this paper, we focus on partitioning transcripts generated by large vocabulary continuous speech recognition (LVCSR), which has achieved great success due to the application of deep neural network (DNN) [19–25].

An important step of story segmentation is to represent the input text data in vector representation, as shown in Fig. 1. A good vector representation is expected to contain story-related information such as content words, while discarding story-irrelevant information such as function words. A commonly used vector representation is bag of words (BOW) [26]. Text (such as a sentence or document) in BOW representation is a high dimensional vector of size $|V| \times 1$, where $|V|$ is the vocabulary size and each element of the vector is the count of a word in the given text. The BOW treats every word equally important, hence function words such as "the" and "a" usually have very high counts, although they are not very useful for story segmentation. A better vector representation for story segmentation is the term frequency-inverse document frequency (tf-idf), which is computed from the BOW and generally assigns high weights to rare content words and low weights to function words. The BOW and tf-idf representations only consider the first order statistics (i.e. counts) of words while ignoring word order and syntactic information that may determine semantic meaning. For example, "the department chair couches offers" and "the chair department offers couches" [27] have very different meanings and belong to different topics, although they have exactly the

* Corresponding author.
E-mail address: lxie@nwpu-aslp.org (L. Xie).

http://dx.doi.org/10.1016/j.sigpro.2017.07.026
0165-1684/

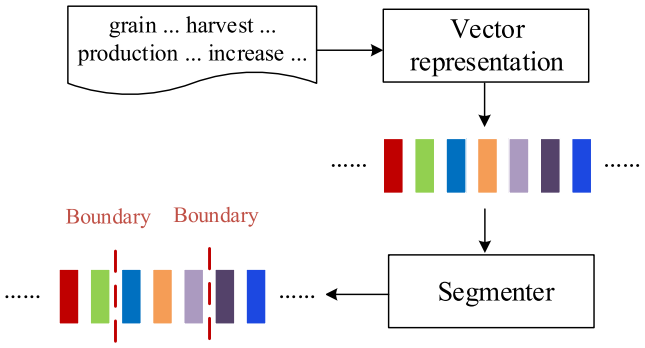

Fig. 1. The diagram of story segmentation process. The words/sentences in the text are first vectorized by tf-idf, BOW, LDA or neural network. Then story segmentation approaches, such as TextTiling, NCuts and HMM, perform partitioning on these vectors.

same BOW representation. Besides, it is difficult to get a high score for synonyms in BOW and tf-idf representations.

In order to mitigate such a limitation, some approaches map the original word to a semantic domain using topic models, such as probabilistic latent semantic analysis (PLSA) [28], latent Dirichlet analysis (LDA) [29] and LapPLSA [28]. These topic models assume that documents are comprised of topics following certain distributions and words are generated from these topics. Significant performance improvements have been observed when the *tf-idf* representation is substituted by a topic representation in both the TextTiling and the normalized cuts (NCuts) approaches [30,31].

Recently, deep learning has achieved great success in various areas due to its strong ability of feature learning and modeling [32–35]. The embedding of words has been explored via various deep learning architectures [32,36,37]. For example, word vector derived from neural network language model (NNLM) can capture syntactic information, such as “quick” and “quickly” , and semantic information such as “Germany” and “Berlin” which is the country to capital city relationship [32]. With these appealing capabilities, word embedding has been successfully applied in various natural language processing (NLP) tasks, including named entity recognition, tagging, and machine translations [36,37]. Since the semantic meaning learned from the neural network is highly task-dependent, in this study, we explore a story segmentation task related topical sentence representation (TSR) through neural network.

Besides vector representation of text, another important step of story segmentation is the segmenter as shown in Fig. 1. Story segmentation methods can be categorized into detection-based methods and probabilistic model-based methods. The former methods find optimal partitions over the word sequence by optimizing local objective criteria, e.g. TextTiling [8,38], or global criteria, e.g. NCuts [39–42]. The probabilistic model based methods assign data with latent random variables (representing topics) and the switch of the latent variable assignments indicates a story boundary. Popular methods in this approach include PLSA [28], BayesSeg [43] and dd-CRP [44].

In this study, we aim to extract a TSR which is optimal for the story segmentation task. Specifically, we are searching for a mapping from a group of words to a topic space, in which vectors from the same topic should group together, while vectors from different topics should be far from each other. As our focus is the learning of vector representation, we use two typical segmenters, i.e. TextTiling [8] and NCuts [41] in this study. Instead of computing lexical similarity by counts of words in the segmenter, we use a neural network to generate vector representations of a sequence of words. The vector representation can be seen as a topic space in which simple Euclidean or cosine of two vectors measures how far the topics of two sequences of words are. We propose three different structures to extract topical sentence representation.

- DNN model: Topical word representation (TSR) extracted from a deep neural network (DNN) trained with long word history. The model maps the local context of a word represented by BOW vectors directly into topic space by predicting the topic class of the input.
- DNN-MTR model: Enhance the DNN model by using BOW vectors of multiple time resolutions (MTR) as input. The long context BOW captures the story information better while the short context BOW provides more detailed information of the current position in the text stream.
- DNN-BNF: Enhance the DNN-MTR model by introducing a bottleneck layer that forces relevant information to be compressed into a relatively low dimensional vector representation, called bottleneck feature (BNF).

We also study the word vector [32] and the sentence vector [45], and compare them with the proposed topical sentence representation. Experiments conducted on the topic detection and tracking (TDT2) corpus [2] show that the proposed topical sentence representation is able to achieve state-of-the-art performance in the story segmentation task.

We summarize our contributions as follows.

- We propose a deep neural network based topical sentence representation specifically for the story segmentation task. By using supervised training, the learned vector representation of sentences contains more topic information and hence more suitable for the story segmentation task.
- We further extend our neural network approach by using multiple time resolution (MTR) information as network input and insertion of a bottleneck (BN) layer. MTR can model both long and short contexts to better capture both slow and fast changing topics. The BN layer in the network results in an information-rich and compact feature representation that leads to improved performance.

The remainder of this paper is organized as follows. Section 2 describes TextTiling and NCuts for story segmentation. Section 3 describes the proposed topical sentence vectors extracted from three different architectures of DNN, and compares them with word/sentence vector. Experiments and analysis are described in Section 4. Finally, we summarize this study and point out our future work in Section 5.

2. Story segmentation approaches

TextTiling [8] and NCuts [38,46] are efficient story segmentation algorithms. These methods detect story boundaries by finding the shift of topics of adjacent sentences or pseudo-sentences. The topic shift itself is detected by computing the similarity between two adjacent sentences, e.g. cosine distance between two vectors that represent the two adjacent sentences. In this section, we briefly review TextTiling and NCuts and leave the vector representation learning and similarity computation to the next section.

2.1. TextTiling

TextTiling [8] has the assumption that different topics employ different sets of words. For example, the words “basketball”, “football” and “baseball” are more likely to appear in a sports news while words “stock market”, “gold price” and “bank” appear frequently in an economic news report. Thus there is a high similarity, called the lexical score, between the sentences in the same story and low similarity between the sentences from different stories. TextTiling locates story boundaries based on the change of