Expert Systems with Applications 42 (2015) 101–112

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Topic identification techniques applied to dynamic language model adaptation for automatic speech recognition

J.D. Echeverry-Correa *, J. Ferreiros-López, A. Coucheiro-Limeres, R. Córdoba, J.M. Montero

Speech Technology Group, E.T.S.I. de Telecomunicación, Av. Complutense, 30. Universidad Politécnica de Madrid, Spain

ARTICLE INFO

Article history:
Available online 10 August 2014

Keywords:
Language model adaptation
Topic identification
Automatic speech recognition

ABSTRACT

In this paper we present an efficient speech recognition approach for multitopic speech by combining information retrieval techniques and topic-based language modeling. Information retrieval based techniques, such as topic identification by means of Latent Semantic Analysis, are used to identify the topic in a recognized transcription of an audio segment. According to the confidence on the topics that have been identified, we propose a dynamic language model adaptation in order to improve the recognition performance in 'a two stages' automatic speech recognition system. The scheme used for the adaptation of the language model is a linear interpolation between a background general LM and a topic dependent LM. We have studied different approaches to generate the topic dependent LM and also for determining the interpolation weight of this model with the background model. In one of these approaches we use the given topic labels in the training dataset to obtain the topic models. In the other approach we separate the documents in the training dataset into topic clusters by using the *k-means* algorithm. For strengthening the adaptation models we also use topic identification techniques to group non topic-labeled documents from the EUROPARL text database in order to increase the amount of data for training specific topic based language models. For the evaluation of the proposed system we are using the Spanish partition of the European Parliament Plenary Sessions (EPPS) Database; we selected a subset of the database with 67 labeled topics for the evaluation. For the task of topic identification our experiments show a relative reduction in topic identification error of 44.94% when compared to the baseline method, the Generalized Vector Model with a classic TF–IDF weighting scheme. For the task of dynamic adaptation of LMs applied to ASR we have achieved a relative reduction in WER of 13.52% over a single background language model.

1. Introduction

The performance of a speech recognition system depends significantly on the similarity between the language model (LM) used and the task that is being addressed. This similarity is even more important in scenarios where the statistical properties of the language fluctuates throughout the time, for example in application domains involving spontaneous and multitopic speech. Over the last years there has been an increasing effort in improving the speech recognition systems for such domains. In spontaneous and multitopic speech the grammar models are changing constantly and therefore the performance of the speech recognition system will depend, among many other parts of the system, on its capacity to update or adapt the LMs. In this paper we propose a dynamic LM adaptation based on an information retrieval (IR) approach. We used IR techniques as a tool for identifying the topic of the speech, thus enabling the system to perform an adaptation of the language model according to the topic that is being discussed.

1.1. Related work in topic identification systems

With the rapid growth of the information available online, topic identification (TI) has become one of the key techniques in the field of text data classification. This technique addresses the problem of identifying which of a set of predefined topics or themes are present in a document. It is currently been applied in many contexts and disciplines, ranging from document indexing to automated metadata generation, document and messages filtering and, in an general sense, in applications that need document separation and organization. Depending on the research discipline in which this task is being carried out, topic identification is also known as *text categorization* (Manning, Raghavan, & Schütze, 2008), *text classification*

* Corresponding author.
E-mail addresses: jdec@die.upm.es (J.D. Echeverry-Correa), jfl@die.upm.es (J. Ferreiros-López), acoucheiro@die.upm.es (A. Coucheiro-Limeres), cordoba@die.upm.es (R. Córdoba), juancho@die.upm.es (J.M. Montero).

http://dx.doi.org/10.1016/j.eswa.2014.07.035
0957-4174/

(Baeza-Yates & Ribeiro-Neto, 2011) or *topic spotting* (Wiener, Pedersen, & Weigend, 1995).

A conventional topic identification framework consists of preprocessing, feature extraction, feature selection and classification stages. The preprocessing stage is usually composed of several tasks such as tokenization, stopword removal, stemming and term categorization. In the feature extraction stage it is common the use of the Vector Space Model (Salton, Yang, & Yu, 1975). This model makes use of the bag of words approach (Baeza-Yates & Ribeiro-Neto, 2011). The feature selection stage generally utilizes filter methods such as weighting schemes for the term and document frequency (Dumais, 1991), techniques for obtaining the mutual information of terms (Liu, Sun, Liu, & Zhang, 2009), information gain (Lee & Lee, 2006) and chi-square statistical metric (Chen & Chen, 2011). The classification stage uses well known techniques from the fields of information retrieval and machine learning systems. In the last years a growing number of statistical learning methods have been applied in TI from these research fields (Sebastiani, 2002). Common approaches includes Latent Semantic Analysis (Deerwester, Dumais, Landauer, Furnas, & Harshman, 1990), Rocchio's method (Rocchio, 1971), Decision Trees (Lewis & Ringuette, 1994), naive Bayesian classifiers (Kim, Rim, Yook, & Lim, 2002) and Support Vector Machines (Joachims, 1998).

TI has been successfully applied in many contexts and disciplines, ranging from topic detection (Qiu, Xu, Li, & Li, 2010), automated metadata generation (Cheng, Chandramouli, & Subbalakshmi, 2011), document and messages filtering (Günal, Ergin, Gülmezoglu, & Gerek, 2006). It has also been applied in recently developed areas, such as sentiment analysis (Maks & Vossen, 2012), genre classification (Petrenz & Webber, 2011) and entity resolution (Pereira et al., 2009), among many other fields of application.

Nevertheless it is interesting to review the influence of TI in the field of language model adaptation. Within this field, TI has been used to study the changes that the language experiences when moving towards different domains (Bellegarda, 2004). In that sense, TI is able to contribute to LM adaptation by adding new sources of information to previously existent models with the objective of enriching them.

1.2. Related work in language model adaptation strategies applied to an ASR system

Over the past 30 years, the amount and diversity of information available online has exponentially grown and this tendency appears to remain unaltered in the near future. As a result, the quality of language models has increased in certain domains where such data became available. Nevertheless, this behavior seems to be reaching an upper limit and it is possible that this continuous increase of information does not lead to any significant improvement in language models (Rosenfeld, 2000). For this reason it is important to find new sources of information that increase the capacity of the data to describe and model the type of language that is being used in an automatic speech recognition application.

For an optimal adaptation of language models in specific domains it is required that the system has a previous knowledge of data belonging to the same domain or, at least, to a related one. Precisely, the aim in statistical language model adaptation is to add new sources of information to the previously existent models with the objective of enriching them. The goal in LM adaptation is to reflect the changes that the language experiences when moving towards different domains or, as in some applications, when dealing with multiple speakers (Bellegarda, 2004).

LM adaptation techniques can be classified according to different criteria. Rosenfeld (2000) proposes a classification based in the domain of the data. Bellegarda (2001), on the other hand, suggests that the classification must be done according to the system requirements. However, there is not a distinct separation between these criteria. Nowadays LM adaptation techniques are jointly based not only on the origin and domain of the data but also on the system requirements and the objective of the adaptation scheme. Some LM adaptation approaches are based on the specific context of the task that they are addressing. In these approaches, new sources of information are used to generate a context-dependent LM which is then merged with a static LM. These new sources of information may come, for instance, from text categorization systems as in Seymore and Rosenfeld (1997), from speaker identification systems (Nanjo & Kawahara, 2003), from linguistic analysis systems (Liu & Liu, 2008) or from the application context itself (Lucas-Cuesta, Ferreiros, Fernández-Martínez, Echeverry, & Lutfi, 2013).

Other approaches are based on analysis and extraction of metadata, which means extraction of information that it is not explicitly described in the text. The topic of a document or semantic information related to it are examples of metadata. Latent Semantic Analysis (LSA) is an example of the type of techniques that exploits this kind of information. The work presented in this paper uses a LSA-based approach in order to obtain information about the topics that are being discussed in an audio segment. This information is then used to dynamically adapt the language model used by an ASR system with the objective of improving the recognition accuracy. Similar works have been proposed in the same domain. In Bellegarda (2000), the use of LSA is proposed to extract the semantic relationships between the terms that appear in a document and the document itself. More robust techniques in the field of information retrieval, as Latent Dirichlet Allocation (LDA) (Blei, Ng, & Jordan, 2003), have also been used for adapting LMs in an automatic speech recognition task (Chien & Chueh, 2011). A keyword extraction strategy to determine the LM to be used in a multi-stage speech recognition system is proposed in (Chen, Gauvain, Lamel, Adda, & Adda, 2001). In contrast to LSA, which do not explicitly consider the exact word order in the history context, in Liu, Gales, and Woodland (2013b) a history weighting function is used to model the change in word history during LM adaptation.

There are also techniques based on information originated from different subsystems or domains (cross adaptation). In Liu, Gales, and Woodland (2013a) a linear combination of two different subsystems (syllable and words) is performed to obtain an adapted LM. Another example is cross-lingual adaptation which uses information in a language to adapt LMs in another language (Kim & Khudanpur, 2004; Tam & Schultz, 2009).

All these techniques have one thing in common and that is the importance of the selection of reliable sources of information for refining the existent models. One of the most common sources of data for adapting language models is the internet. When using data available online it is possible to find information related to a large variety of topics. Nevertheless, this broad coverage leads to a loss of specificity when estimating LMs (Lucas-Cuesta et al., 2013). To avoid this drawback, clustering algorithms have been proposed to group together those elements that share some properties. Topic-based language modeling is an example of this clustering criterion (Chen, Seymore, & Rosenfeld, 1998; Iyer & Ostendorf, 1999).

Techniques, in the line of Latent Semantic Analysis (Deerwester et al., 1990) such as Probabilistic Latent Semantic Analysis (PLSA) (Hofmann, 1999) and Latent Dirichlet Allocation have been proposed to group documents into topic clusters. Topic based language models can be found in a broad spectrum of applications, such as in information retrieval systems as part of the ranking function (Zhai, 2008), in spoken dialogue systems for adapting the speech recognizer to the dialogue context (López-Cózar & Callejas, 2006; Lucas-Cuesta, 2013), in dynamic language model adaptation for Large Vocabulary Continuous Speech Recognition