AUTHOR-TOPIC BASED REPRESENTATION OF CALL-CENTER CONVERSATIONS

Mohamed Morchid, Richard Dufour, Mohamed Bouallegue, Georges Linarès

LIA - University of Avignon (France)
{firstname.lastname}@univ-avignon.fr

ABSTRACT

Performance of Automatic Speech Recognition (ASR) systems drops dramatically when transcribing conversations recorded in noisy conditions. Speech analytics suffer from this poor automatic transcription quality. To tackle this difficulty, a solution consists in mapping transcriptions into a space of hidden topics. This abstract representation allows to substantiate the drawbacks of the ASR process. The well-known and commonly used one is the topic-based representation from a Latent Dirichlet Allocation (LDA). Several studies demonstrate the effectiveness and reliability of this high-level representation. During the LDA learning process, distribution of words into each topic is estimated automatically. Nonetheless, in the context of a classification task, no consideration is made for the targeted classes. Thus, if the targeted application is to find out the main theme related to a dialogue, this information should be taken into consideration. In this paper, we propose to compare a classical topic-based representation of a dialogue, with a new one based not only on the dialogue content itself (words), but also on the theme related to the dialogue. This original representation is based on the author-topic (AT) model. The effectiveness of the proposed representation is evaluated on a classification task from automatic dialogue transcriptions between an agent and a customer of the Paris Transportation Company. Experiments confirmed that this author-topic model approach outperforms by far the classical topic representation, with a substantial gain of more than 7% in terms of correctly labeled conversations.

Index Terms— Author-Topic model, Human/human conversation, Speech recognition, Latent Dirichlet Allocation, Classification

1. INTRODUCTION

Automatic Speech Recognition (ASR) systems frequently fail on noisy conditions and high Word Error Rates (WERs) make difficult the analysis of the automatic transcriptions. Applications suffer from these transcription issues that may be overcome by improving the ASR robustness or/and the tolerance of speech analytic systems to ASR errors. Telephone conversation is a particular case of human/human interaction where automatic processing encounters many difficulties, especially due to the speech recognition step required to transcribe the speech contents. First, the speaker behavior may be unexpected and the train/test mismatch may be very large. Second, speech signal may be strongly impacted by various sources of variability: environment and channel noises, acquisition devices...

One purpose of the telephone conversation application is to identify the main theme that appears in the conversation. Themes are related to the reason why the customer called. In this application, 8 classes corresponding to the main customer requests are considered (*lost and founds, traffic state, timelines...*). Additionally to the classical problems in such adverse conditions, the topic identification system should also face issues due to class proximity. For example, a *lost & found* request is related to itinerary (*where the object was lost?*) or timeline (*when?*), that could appear in most of the classes. In fact, these conversations involve a relatively small set of basic concepts related to transportation issues.

An efficient way to tackle both ASR robustness and class ambiguity is to map dialogues into a topic space abstracting the ASR outputs. Then, dialogues classification will be achieved in this topic space. Numerous unsupervised methods to estimate topic-spaces were proposed in the past. Latent Dirichlet Allocation (LDA) [1] was largely used in speech analytics applications [2]. During the LDA learning process, distribution of words into each topic is estimated automatically. Nonetheless, the class associated to the dialogue is not directly taken into account in the topic model. Indeed, the classes are usually only used to train a classifier at the end of the process. As a result, such a system considers separately the document content (*i.e.* words), to learn a topic model, and the labels (*i.e.* classes) to train a classifier. We can however note that, in the considered application, dialogues are labeled by a human annotator: a relation between the document content (words) and the document label (class) should then exist.

In the context of dialogue classification, this relation is crucial to efficiently label unseen (*i.e.* new) dialogues. This model (LDA) needs to infer an unseen document into each topic space. The processing time during the inference pro-

This work was funded by the SUMACC and ContNomina projects supported by the French National Research Agency (ANR) under contracts ANR-10-CORD-007 and ANR-12-BS02-0009.

cess as well as the difficult choice of an efficient number of iterations, do not allow us to evaluate effectively and quickly the best theme related to a given document. In this paper, we propose to build a topic model, called author-topic (AT) model [3], that takes into consideration all information contained into a document: the content itself (*i.e.* words), the label (*i.e.* class) and the relation between the distribution of words into the document and the label, considered as a latent relation. From this model, a vector representation in a continuous space is built for each dialogue. Then, a supervised classification approach, based on Support Vector Machines (SVM) [4], is applied. This method is evaluated in the application framework of the RATP call centre (Paris Public Transportation Authority), focusing on the theme identification task [5].

The rest of this paper is organized as follows: Topic model representations from document content are described in Section 2, by introducing LDA and AT models. Section 3 presents the experimental protocol while Section 4 reports classification results. Finally, Section 5 concludes the work and gives some perspectives.

2. TOPIC-MODELING FOR AUTOMATIC TRANSCRIPTIONS

Dialogues, automatically transcribed using an Automatic Speech Recognition (ASR) system, contain many errors due to noisy recording conditions. An elegant way to tackle these errors is to map dialogues in a thematic space in order to abstract the document content. The most known and used one is the Latent Dirichlet Allocation (LDA) [1] model. The LDA approach represents documents as a mixture of latent topics. Nonetheless, this model does not code statistical relations between words contained into the document, and the label that could be associated to it.

To go beyond this limit, the Author-topic (AT) model [3] has been proposed. The AT model links both authors (here, the label) and documents content (words). The next sections describe both LDA and AT models. Examples of a dialogue mapped into a topic space from LDA and AT models are presented respectively in Figures 1 and 3.

2.1. Latent Dirichlet Allocation (LDA)

Several approaches proposed to consider the document as a mixture of latent topics. These methods, such as Latent Semantic Analysis (LSA) [6, 7], Probabilistic LSA (PLSA) [8] or Latent Dirichlet Allocation (LDA) [1], build a higher-level representation of the document in a topic space. Documents are then considered as a *bag-of-words* [9] where the word order is not taken into account.

LDA is presented into its plate notation in Figure 2 (a). These methods demonstrated their performance on various tasks, such as sentence [10] or keyword [11] extraction. In opposition to a multinomial mixture model, LDA considers that a theme is associated to each occurrence of a word composing the document, rather than associate a topic with the complete document. Thereby, a document can change of topics from a word to another. However, the word occurrences are connected by a latent variable which controls the global respect of the distribution of the topics in the document. These latent topics are characterized by a distribution of word probabilities which are associated with them. PLSA and LDA models have been shown to generally outperform LSA on IR tasks [12]. Moreover, LDA provides a direct estimate of the relevance of a topic knowing a word set.

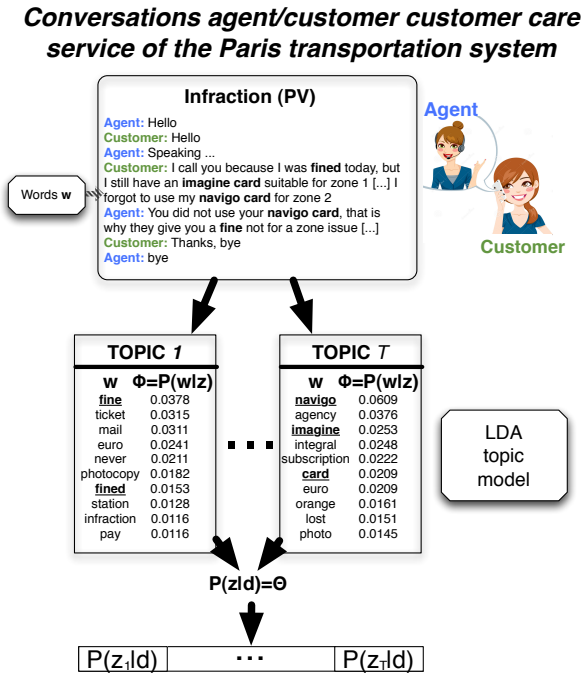


Fig. 1. Example of a dialogue d mapped into a topic space of size n.

The generative process corresponds to the hierarchical Bayesian model shown, using plate notation, in Figure 2 (a). Several techniques, such as Variational Methods [1], Expectation-propagation [13] or Gibbs Sampling [14], have been proposed to estimate the parameters describing a LDA hidden space. Gibbs Sampling is a special case of Markov-chain Monte Carlo (MCMC) [15] and gives a simple algorithm for approximate inference in high-dimensional models such as LDA [16]. This overcomes the difficulty to directly and exactly estimate parameters that maximize the likelihood of the whole data collection defined as:

$$P(W|\overrightarrow{\alpha},\overrightarrow{\beta}) = \prod_{w \in W} P(\overrightarrow{w}|\overrightarrow{\alpha},\overrightarrow{\beta}) \qquad (1)$$

for the whole data collection W knowing the Dirichlet param-

eters $\overrightarrow{\alpha}$ and $\overrightarrow{\beta}$.

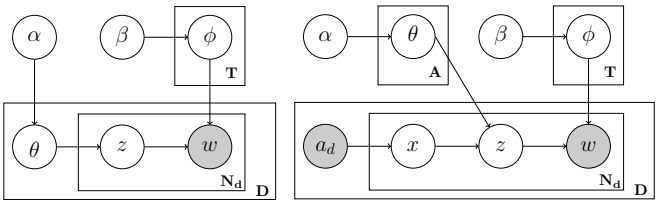
The Gibbs Sampling, to estimate LDA, was firstly reported in [14]. A more comprehensive description of this method can be found in [16]. One can refer to these papers for a better understanding of this sampling technique. This method is used both to estimate the LDA parameters and to infer an unseen dialogue with a hidden space of T topics.

In the LDA technique, the topic z is drawn from a multinomial over θ which is drawn from a Dirichlet($\overrightarrow{\alpha}$). Thus, a set of p topic spaces are learned using LDA by varying the number of topics T to obtain p topic spaces of size T.

Gibbs Sampling allows one to both estimate the LDA parameters, in order to represent a new dialogue d with the r^{th} topic space of size T, and obtain a feature vector $V_d^{z^r}$ of the topic representation of d. The j^{th} feature:

$$V_d^{z_j^r} = \theta_{j,d}^r \,, \tag{2}$$

where $\theta_{j,d}^r = P(z_j^r|d)$ is the probability of topic z_j^r ($1 \leq j \leq T$) generated by the unseen dialogue d in the r^{th} topic space of size T as described in Figure 1.



(a) LDA (b) Author-topic model

Fig. 2. Generative models in plate notation for Latent Dirichlet Allocation (LDA) (a) and Author-Topic (AT) (b) models.

2.2. Author-topic (AT) model

The Author-topic (AT) model, represented into its plate notation in Figure 2 (b), uses a topic-based representation to model both the document content (words distribution) and the authors (authors distribution). For each word w contained into a document d, an author a is uniformly chosen at random. Then, a topic z is chosen from a distribution over topics specific to that author, and the word is generated from the chosen topic.

In our considered application, a document d is a conversation between an agent and a customer. The agent have to label this dialogue with one of the 8 defined themes, a theme being considered as an author. Thus, each dialogue d is composed with a set of words w and a theme a. In this model, x indicates the author (*i.e.* the theme) responsible for a given word, chosen from a_d. Each author is associated with a distribution over topics (θ), chosen from a symmetric Dirichlet prior ($\overrightarrow{\alpha}$), and a weighted mixture to select a topic z. A word is then generated according to the distribution ϕ corresponding to the topic z. This distribution ϕ is drawn from a Dirichlet ($\overrightarrow{\beta}$).

The parameters ϕ and θ are estimated from a straightforward algorithm based on Gibbs Sampling such as LDA hyperparameters estimation method (see Section 2.1). One can find more about Gibbs Sampling and author topic model in [3].

Figure 3 shows the mapping process of an unseen dialogue d from the validation set, into an author topic space of size T. Each dialogue d is composed with a set of words w and a label (or theme) a considered as the author in the AT model. Thus, this model allows one to code statistical dependencies between dialogue content (words w) and label (theme a) through the distribution of the latent topics into the dialogue.

Gibbs Sampling allows us to estimate the AT model parameters, in order to represent an unseen dialogue d with the r^{th} author topic space of size T, and to obtain a feature vector $V_d^{a_k^r} = P(a_k|d)$ of the topic representation an unseen dialogue d with the r^{th} author topic space Δ_r^n of size T. The k^{th} ($1 \leq k \leq A$) feature is:

$$V_d^{a_k^r} = \sum_{i=1}^{N_d} \sum_{j=1}^{T} \theta_{j,a_k}^r \phi_{j,i}^r \tag{3}$$

where A is the number of authors (or themes); $\theta_{j,a_k}^r = P(a_k|z_j^r)$ is the probability of author a_k to be generated by the topic z_j^r ($1 \leq j \leq T$) in the r^{th} topic space of size T. $\phi_{j,i}^r = P(w_i|z_j^r)$ is the probability of the word w_i (N_d is the vocabulary size of d) to be generated by the topic z_j^r.

3. EXPERIMENTAL PROTOCOL

We propose to evaluate the effectiveness of the proposed approach in the application framework of the DECODA corpus [5]. This corpus is composed of a set of dialogues between agents and customers. A classification approach based on Support Vector Machines (SVM) is performed to find out the main theme of a given dialogue. The next sections describe the dataset as well as the classification method.

3.1. Dialogues dataset

The DECODA project corpus [5] was used to perform experiments on the conversation theme identification. It is composed of 1,514 telephone conversations, corresponding to about 74 hours of signal, split into a train set (740 dialogues), a development set (175 dialogues) and a test set (327 dialogues), and manually annotated with 8 ($A = 8$) conversation themes (or authors a in the author topic model): *problems of itinerary*, *lost and found*, *time schedules*, *transportation cards*, *state of the traffic*, *fares*, *infractions* and *special offers*.

38 topic spaces are elaborated by varying the number of topics from 10 to 200 (step of 5 topics). The topic spaces are

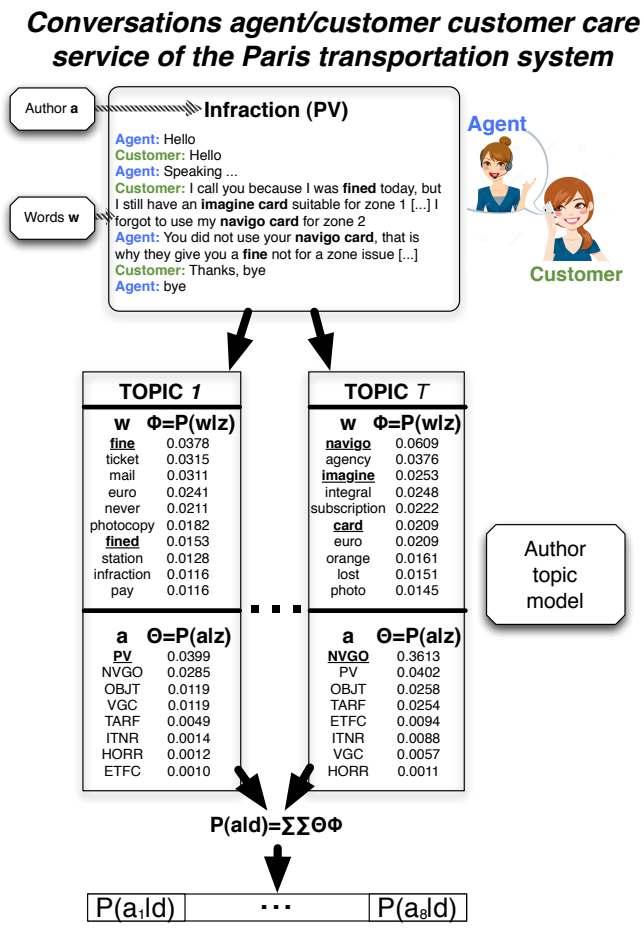


Fig. 3. Example of a dialogue d mapped into an author topic model of size n.

learned with an homemade implementation of LDA and AT models.

The LIA-Speeral ASR system [17] has been used for the experiments. Acoustic model parameters were estimated from 150 hours of speech in telephone conditions. The vocabulary contains 5,782 words. A 3-gram language model (LM) was obtained by adapting a basic LM with the train set transcriptions. This system reaches an overall Word Error Rate (WER) of 45.8%, 59.3%, and 58.0%, respectively on the train, development and on test sets. These high WER are mainly due to speech disfluencies and to adverse acoustic environments (for example, calls from noisy streets with mobile phones). A "stop list" of 126 words[1] was used to remove unnecessary words (mainly function words) which results in a WER of 33.8% on the train, 45.2% on the development, and 49.5% on the test.

For sake of comparison, experiments are conducted using the manual transcriptions only (TRS) and the automatic transcriptions only (ASR). The conditions indicated by the abbreviations between parentheses are considered for the development (Dev) and the test (Test) sets.

Only homogenous conditions (TRS or ASR for both training and validations sets) are considered in this study. Authors in [2] notice that results collapse dramatically when heterogenous conditions are employed (TRS or TRS+ASR for training set and ASR for validation set).

3.2. SVM classification

As the classification of dialogues requires a multi-class classifier, the SVM *one-against-one* method is chosen with a linear kernel. This method gives a better accuracy than the *one-against-rest* [18]. In this multi-theme problem, A denotes the number of themes and $t_i, i = 1, \ldots, A$ denotes the A themes. A binary classifier is used with a linear kernel for every pair of distinct theme. As a result, binary classifiers $A(A-1)/2$ are constructed all together. The binary classifier $C_{i,j}$ is trained from example data where t_i is a positive class and t_j a negative one ($i \neq j$).

For a vector representation of an unseen dialogue d ($V_d^{z_j^r}$ for a LDA representation and $V_d^{a_k^r}$ for an AT representation), if $C_{i,j}$ means that d is in the theme t_i, then the vote for the class t_i is added by one. Otherwise, the vote for the theme t_j is increased by one. After the vote of all classifiers, the dialogue d is assigned to the theme having the highest number of votes.

Table 1. Theme classification accuracy (%) best configuration from development set applied to test set.

Topic Model	DATASET		Best Dev		Test
	Train	Test	#topics	acc. %	acc. %
LDA	TRS	TRS	150	89.7	82.5
LDA	ASR	ASR	160	82.3	73.1
AT	TRS	TRS	175	89.7	**83.7**
AT	ASR	ASR	45	89.1	**80.4**

4. RESULTS

The results obtained using manual (TRS) and automatic (ASR) transcriptions with respectively a topic-based representation from LDA and from an AT model, are shown in Figures 4 and 5. We can firstly point out that, for both representations of dialogues (LDA or AT), the best results are obtained with manual transcriptions (TRS), which correspond to classical textual documents. This could be explained by the fact that the document content (words) does not suffer from ASR errors. We can see that the AT representation outperforms LDA, no matter the corpus (development or test) or the conditions (TRS/ASR) studied.

[1] http://code.google.com/p/stop-words/

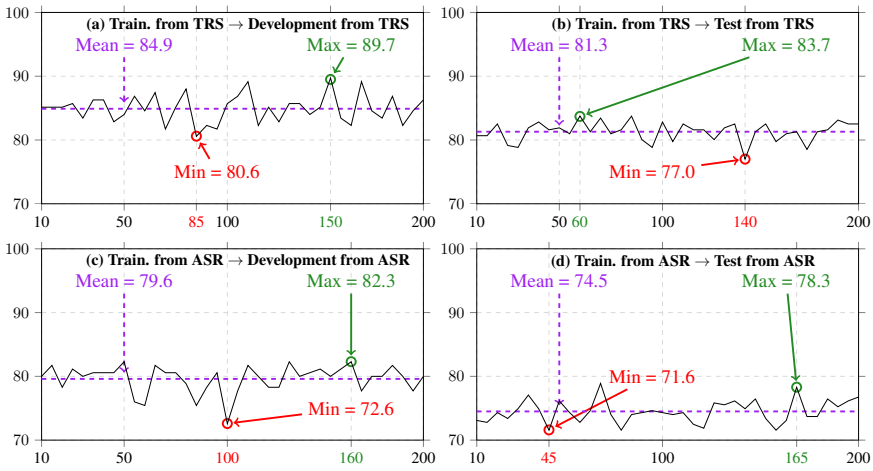


Fig. 4. Theme classification accuracies (%) using various LDA topic-based representations on the development and test sets with different experimental configurations. X-axis represents the number n of classes contained into the topic space ($10 \leq n \leq 200$).

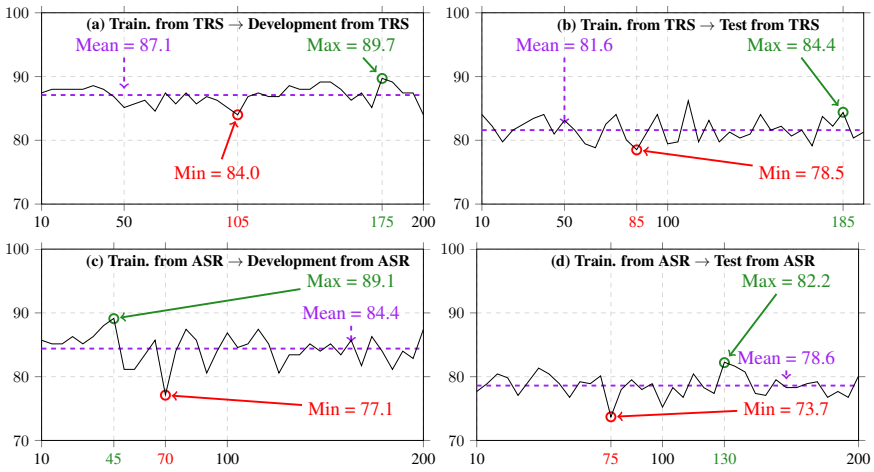


Fig. 5. Theme classification accuracies (%) using various author topic-based representations on the development and test sets with different experimental configurations. X-axis represents the number n of classes contained into the topic space ($10 \leq n \leq 200$).

The first and the most intuitive experiment made, is to evaluate the best theme a, found from the AT model ($V_d^{a_k^r} = P(a_k|d)$), which maximizes $P(a_k|d)$. The results obtained with this simple evaluation are quite low (less than 44% of accuracy). For this reason, these results are not reported here.

In order to better compare performance obtained by both approaches (LDA/AT), best results are reported in Table 1. Note that these results are given in "real" application condition, *i.e.* the best configuration (number of topics contained into the topic space) being chosen with the development set. As a result, a better operating point could exist in the test set, which could explain the performance difference between results reported in Table 1, and Figures 4 and 5. With this real condition, we can note that the AT model allows to outperform the LDA approach, with a gain of 1.2 points using the manual transcriptions (TRS) and of 7.3 points using the automatic transcriptions (ASR).

Another interesting point, is the stability and robustness of the AT model curve of the development set in TRS condition, comparatively to the LDA representation. Indeed, the results are mainly close to the mean value (87.1%). The maximum achieved by both representations in TRS condition are the same. Thus, knowing that dialogues are labeled (annotated) by an agent, and the fact that a dialogue may contain more than only one theme, this maximum represents the limit of a topic-based representation in a multi-theme context. Nonetheless, this remark is not applicable to the ASR condition.

5. CONCLUSION

Performance of ASR systems depends strongly to the recording environment. In this paper, an elegant way to deal with ASR errors by mapping a dialogue into an Author-topic (AT) space is presented. This high-level representation allows us to significantly improve the performance of the theme identification task. Experiments conducted on the DECODA corpus showed the effectiveness of the proposed AT model in comparison to the use of the classic LDA representation, with a gain of more than 1 and 7 points respectively using manual and automatic transcriptions.

In the future, an interesting work will be to compare this representation with other ones using a thematic representation, such as labeled LDA [19] or supervised LDA [20], since this last one is close to the author topic representation. Another perspective is to add a new latent variable into the author topic model, to allow this model to infer effectively an unseen dialogue.

6. REFERENCES

[1] David M. Blei, Andrew Y. Ng, and Michael I. Jordan, "Latent dirichlet allocation," *The Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[2] Mohamed Morchid, Richard Dufour, Pierre-Michel Bousquet, Mohamed Bouallegue, Georges Linarès, and Renato De Mori, "Improving dialogue classification using a topic space representation and a gaussian classifier based on the decision rule," in *ICASSP*, 2014.

[3] M. Rosen-Zvi, T. Griffiths, M. Steyvers, and P. Smyth, "The author-topic model for authors and documents," in *Proceedings of the 20th conference on Uncertainty in artificial intelligence*. AUAI Press, 2004, pp. 487–494.

[4] Vladimir Vapnik, "Pattern recognition using generalized portrait method," *Automation and Remote Control*, vol. 24, pp. 774–780, 1963.

[5] Frederic Bechet, Benjamin Maza, Nicolas Bigouroux, Thierry Bazillon, Marc El-Beze, Renato De Mori, and Eric Arbillot, "Decoda: a call-centre human-human spoken conversation corpus," LREC'12, 2012.

[6] S. Deerwester, S.T. Dumais, G.W. Furnas, T.K. Landauer, and R. Harshman, "Indexing by latent semantic analysis," *Journal of the American society for information science*, vol. 41, no. 6, pp. 391–407, 1990.

[7] Jerome R. Bellegarda, "A latent semantic analysis framework for large-span language modeling," in *Fifth European Conference on Speech Communication and Technology*, 1997.

[8] Thomas Hofmann, "Probabilistic latent semantic analysis," in *Proc. of Uncertainty in Artificial Intelligence, UAI ' 99*. Citeseer, 1999, p. 21.

[9] Gerard Salton, "Automatic text processing: the transformation," *Analysis and Retrieval of Information by Computer*, 1989.

[10] Jerome R. Bellegarda, "Exploiting latent semantic information in statistical language modeling," *Proceedings of the IEEE*, vol. 88, no. 8, pp. 1279–1296, 2000.

[11] Yoshimi Suzuki, Fumiyo Fukumoto, and Yoshihiro Sekiguchi, "Keyword extraction using term-domain interdependence for dictation of radio news," in *17th international conference on Computational linguistics*. ACL, 1998, vol. 2, pp. 1272–1276.

[12] Thomas Hofmann, "Unsupervised learning by probabilistic latent semantic analysis," *Machine Learning*, vol. 42, no. 1, pp. 177–196, 2001.

[13] Thomas Minka and John Lafferty, "Expectation-propagation for the generative aspect model," in *Conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 2002, pp. 352–359.

[14] Thomas L. Griffiths and Mark Steyvers, "Finding scientific topics," *Proceedings of the National academy of Sciences of the United States of America*, vol. 101, no. Suppl 1, pp. 5228–5235, 2004.

[15] Stuart Geman and Donald Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, , no. 6, pp. 721–741, 1984.

[16] Gregor Heinrich, "Parameter estimation for text analysis," *Web: http://www. arbylon. net/publications/text-est. pdf*, 2005.

[17] Georges Linarès, Pascal Nocéra, Dominique Massonie, and Driss Matrouf, "The lia speech recognition system: from 10xrt to 1xrt," in *Text, Speech and Dialogue*. Springer, 2007, pp. 302–308.

[18] Guo-Xun Yuan, Chia-Hua Ho, and Chih-Jen Lin, "Recent advances of large-scale linear classification," vol. 100, no. 9, pp. 2584–2603, 2012.

[19] Daniel Ramage, David Hall, Ramesh Nallapati, and Christopher D. Manning, "Labeled lda: A supervised topic model for credit attribution in multi-labeled corpora," in *Conference on Empirical Methods in Natural Language Processing*, 2009, pp. 248–256.

[20] Jon D. Mcauliffe and David M. Blei, "Supervised topic models," in *Advances in neural information processing systems*, 2008, pp. 121–128.