IMPROVING DIALOGUE CLASSIFICATION USING A TOPIC SPACE REPRESENTATION AND A GAUSSIAN CLASSIFIER BASED ON THE DECISION RULE

Mohamed Morchid[†], *Richard Dufour*[†], *Pierre-Michel Bousquet*[†], *Mohamed Bouallegue*[†]
Georges Linarès[†] *and Renato De Mori*[†‡], *Fellow, IEEE*

[†]LIA, University of Avignon (France)
[‡]McGill University, School of Computer Science, Montreal, Quebec (Canada)
{firstname.lastname}@univ-avignon.fr, rdemori@cs.mcgill.ca

ABSTRACT

In this paper, we study the impact of dialogue representations and classification methods in the task of theme identification of telephone conversation services having highly imperfect automatic transcriptions. Two dialogue representations are firstly compared: the classical Term Frequency-Inverse Document Frequency with Gini purity criteria (TF-IDF-Gini) method and the Latent Dirichlet Allocation (LDA) approach. We then propose to study an original classification method that takes advantage of the LDA topic space representation, highlighted as the best dialogue representation. To do so, two assumptions about topic representation led us to choose a Gaussian process (GP) based method. This approach is compared with a Support Vector Machine (SVM) classification method. Results show that the GP approach is a better solution to deal with the multiple theme complexity of a dialogue, no matter the conditions studied (manual or automatic transcriptions). We finally discuss the impact of the topic space reduction on the classification accuracy.

Index Terms— Speech analytics, Theme classification, Latent dirichlet allocation, SVM, Gaussian process

1. INTRODUCTION

The automatic analysis of telephone conversations is a particular case of human/human interactions that involves many difficulties. Indeed, the customer behavior is highly unpredictable which leads to conversations that may contain very noisy segments. When transcribed by an Automatic Speech Recognition (ASR) system, these highly imperfect transcriptions are difficult to exploit.

One purpose of the telephone conversation application is to identify themes that appear in the conversation. A conversation may contain more than one semantically related theme, some of them being irrelevant for the application task. Agents then annotate a conversation with what they consider the major theme of the customer request: as a result, a single theme is associated for each conversation.

In this paper, we firstly investigate the use of two dialogue representation methods by firstly comparing their performance with the same classification algorithm. We then propose to evaluate two classification methods using the best chosen dialogue representation.

In the context of Information Retrieval (IR) tasks, the main feature used is the *term frequency* that allows to obtain a subset of discriminative[1] words for a considered class. This set of discriminative words should permit to compose a vector representation of conversation themes in the semantic space. Its application to automatic transcriptions is more difficult since transcription errors would lead to an incorrect word representation. Thereby, we assume that dialogues have to be considered in an intermediate thematic representation to fully perform this multiple theme complexity. For this reason, the projection of the automatically transcribed words in a more abstracted space could increase the robustness to the ASR errors.

Thus, we propose to explore a term frequency, with the TF-IDF-GINI method, and a topic space representation, with a Latent Dirichlet Allocation (LDA) approach [1], coupled with a classification method to automatically identify themes from highly imperfect transcriptions.

The other main issue is the choice of the best classification method that does not modify the dialogue topic representation. In the second part of this paper, the classical SVM method [2], that modifies the dialogue representation with a kernel function, is compared with a Naive Bayesian classifier, that does not modify it. We assume that this original study will highlight the fact that these two assumptions are relevant: the Gaussianity of the theme classes and the equality of the class covariances.

We finally discuss the impact of the LDA topic space granularity and the space reduction on the theme classification accuracy. In particular, we want to show that the impact of this space reduction varies depending on the number of topics considered.

The paper is organized as follows. Section 2 presents the related work. The dialogue representation approaches and the classification methods are described in sections 3 and 4. Sections 5 and 6 reports experimental results before concluding in section 7.

2. RELATED WORK

The classical Term Frequency-Inverse Document Frequency (TF-IDF) [3] has been widely used for extracting discriminative words. Improvements are observed with the Gini purity criteria [4].

Other approaches proposed to consider the document as a mixture of latent topics. These methods, such as Latent Semantic Analysis (LSA) [5, 6], Probabilistic LSA (PLSA) [7] 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 [8] where the word order is not taken into account. PLSA and LDA models generally outperform LSA on IR tasks [9].

This work was funded by the SUMACC and DECODA projects supported by the French National Research Agency (ANR) under contracts ANR-10-CORD-007 and ANR-09-CORD-005.

[1]The term "discriminative" is associated to a word if it permits to discern a class from the others.

Various classification approaches have been studied. One of the most used is the Support Vector Machine (SVM) method. SVMs are a set of supervised learning techniques. Knowing a sample, SVMs determine a separation plan between parts of the samples called *support vector*. Then, a separating hyperplane that maximizes the *margin* between the support vectors and the hyperplane separator [10] is calculated. SVMs were used for the first time by [11] both in regression [12] and in classification [13] tasks.

A LDA-based approach combined with a SVM classification process has recently been studied in various domains, such as biology [14], text classification [15], audio information retrieval [16], social event detection [17] or image detection [18]. A combined LDA-SVM approach has been explored in the context of keyword extraction in automatic transcriptions [19], but not in the context of the theme classification of highly imperfect automatic transcriptions.

The Gaussian classifier based on a Bayes decision rule has been studied mainly in speaker identification from audio, such as [20], where the authors use a compact version of a Gaussian Mixture Model (GMM) super-vector (named *i-vector*), or in [21], where a within covariance matrix of normalized data to represent the intersession variability is proposed. The Mahalanobis [22] metric distance is generally used to evaluate this particular task. To our knowledge, a combined LDA-Gaussian-based Bayes approach has not yet been applied for this particular multi-theme classification problem.

3. DIALOGUE REPRESENTATION

The next sections describe two different unsupervised approaches to create a vector representation of words: a term frequency Okapi/BM25 vector [3] with the TF-IDF-Gini method [4] and a topic space representation with the LDA approach [1].

3.1. Term frequency representation using discriminative terms

Let's consider a corpus D of dialogues d with a word vocabulary $\mathbf{V} = \{w_m\}_{m=1}^{N}$ of size N where d is seen as a bag-of-words [8]. A term w of $\mathbf{V}$ is chosen from its importance $\delta_t^w = p(w|t) = tf(w)idf(w)gini(w)$ in the theme t. $gini(w)$ is common for all the themes. Then the words having the highest scores Δ for all the themes $\mathbf{T}$ constitute a discriminative word subset $\mathbf{V_\Delta}$. Each theme $t \in \mathbf{T}$ has its own score δ_t and its own frequency $\gamma_t = p(t)$ which is the frequency of the dialogues $d \in t$ in the corpus D. Note that a same word w can be present in different themes, but with different scores depending of its relevance in the theme:

$$\begin{aligned}\Delta(w) = p(w|t, t \in \mathbf{T}) &= \sum_{t \in \mathbf{T}} p(w|t)p(t) \\ &= \left\langle \overrightarrow{\delta^w}, \overrightarrow{\gamma} \right\rangle_{t \in \mathbf{T}} . \end{aligned} \quad (1)$$

For each dialogue $d \in D$, a semantic feature vector V_d^s is determined. The n^{th} $(1 \leq n \leq |\mathbf{V_\Delta}|)$ feature $V_d^s[n]$ is composed with the number of occurrences of the word w_n ($|w_n|$) in d and the score Δ of w_n (see equation 1) in the discriminative word set $\mathbf{V_\Delta}$ defined as $V_d^s[n] = |w_n| \times \Delta(w_n)$.

3.2. Topic representation

The topic representation is performed using a Latent Dirichlet Allocation (LDA) approach. The LDA parameters are estimated by using the Gibbs sampling technique. This is due to the difficulty to directly and exactly estimate parameters that maximize the likelihood of the whole data collection defined as: $p(W|\overrightarrow{\alpha}, \overrightarrow{\beta}) = \prod_{m=1}^{M} p(\overrightarrow{w}_m|\overrightarrow{\alpha}, \overrightarrow{\beta})$ for the whole data collection $W = \{\overrightarrow{w}_m\}_{m=1}^{M}$ knowing the Dirichlet parameters $\overrightarrow{\alpha}$ and $\overrightarrow{\beta}$.

Several techniques to estimate the LDA parameters exist, such as Variational Methods [1], Expectation-propagation [23] or Gibbs sampling [24]. Gibbs sampling is a special case of Markov-chain Monte Carlo (MCMC) [25] and gives a simple algorithm to approximate inference in high-dimensional models such as LDA [26]. The first use of Gibbs sampling for estimating LDA is reported in [24] and a more comprehensive description of this method is available in the technical report [26]. This method is used both to estimate parameters of LDA and to infer an unseen dialogue with the topic space. Thus, the Gibbs sampling allows to obtain a feature vector V_d^z of the topic representation of d. The features considered for describing a dialogue d are the probabilities $p(z_f|d)$, where z_n $1 \leq f \leq n$ is a *hidden topic* belonging to a hidden topic space. Let V_d^z be the vector of these probabilities.

4. CLASSIFICATION METHODS

This section presents the proposed theme classification approaches that use the extracted vectors V_d^z to learn a classifier (SVM or Gaussian-based approaches).

4.1. Gaussian-based Bayes classifier

This probabilistic approach ignores the process by which vectors were extracted. Instead, they pretend they were generated by a prescribed generative model. Once a topic vector is obtained from a dialogue, the LDA mechanism is ignored and is considered as an observation from a probabilistic generative model. The two most simple assumptions are those of the homoscedastic Gaussian-based Bayes classifier [27]: (i) the Gaussianity of the theme classes and (ii) the equality of the class covariances.

The Gaussian classifier is based on the Bayes decision rule and is combined with a scoring metric to assign the most likely theme $\hat{t}$ to a dialogue d. Given a training dataset D of dialogues, let $\mathbf{W}$ denote the within dialogue covariance matrix defined by:

$$\mathbf{W} = \sum_{k=1}^{K} \frac{n_t}{n} \mathbf{W_k} = \frac{1}{n} \sum_{k=1}^{K} \sum_{i=0}^{n_t} \left(x_k^i - \overline{x_k}\right) \left(x_k^i - \overline{x_k}\right)^t \quad (2)$$

where K is the number of themes, $\mathbf{W_k}$ is the covariance matrix of the k^{th} theme C_k, n_t is the number of dialogues annotated with for the theme t_k, n is the total number of dialogues in the training set, x_k^i is the vector of features for the i^{th} dialogue annotated with the k^{th} theme and $\overline{x_k}$ is the centroid of all vectors x_k^i describing the features of dialogues annotated with the k^{th} theme. Each dialogue does not contribute to the covariance in an equivalent way. For this reason, the term $\frac{n_t}{n}$ is introduced in equation 2.

If homoscedasticity (equality of the class covariances) and Gaussian conditional density models are assumed, a new observation x from the test dataset can be assigned to the most likely theme k_{Bayes} using the Gaussian classifier based on the Bayes decision rule:

$$\hat{t}_{\text{Bayes}} = \arg\max_k \left\{ -\frac{1}{2} \left(x - \overline{x_k}\right)^t \mathbf{W}^{-1} \left(x - \overline{x_k}\right) + a_k \right\} \quad (3)$$

where x is the feature vector of a document d, $\mathbf{W}$ is the within theme covariance matrix defined in equation 2, $\mathcal{N}$ denotes the normal distribution and a_k is the log prior probability of the theme membership ($a_k = \log(p(C_k))$). It is worth noting that, with these

assumptions, the Bayesian approach is similar to the Fisher's geometric approach: x is assigned to the nearest centroid's class, according to the Mahalanobis [22] metric of $\mathbf{W}^{-1}$:

$$\hat{t}_{\text{Bayes}} = \arg\max_k \left\{ -\frac{1}{2}||x - \overline{x_k}||^2_{\mathbf{W}^{-1}} + a_k \right\} \quad (4)$$

4.2. SVM classification

This classifier modifies the representation of dialogues, which are mapped into a space of higher dimension. This goes against the assumptions previously defined in section 4.1.

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* [2]. In this multi-theme problem, T denotes the number of themes and $t_i, i = 1, \ldots, T$ denotes the T themes. A binary classifier is used with a linear kernel for every pair of distinct theme. As a result, binary classifiers $T(T-1)/2$ 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, 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.

5. EXPERIMENTAL PROTOCOL

Experiments are performed using The DECODA project corpus [28]. This corpus is composed of 1,514 telephone conversations split into a train set (740 dialogues), a development set (447 dialogues) and a test set (327 dialogues), and manually annotated with 8 conversation themes: *problems of itinerary*, *lost and found*, *time schedules*, *transportation cards*, *state of the traffic*, *fares*, *infractions* and *special offers*.

The ASR system used for the experiments is the LIA-Speeral system [29]. Model parameters were estimated with maximum a-posteriori probability (MAP) adaptation from 150 hours of speech in telephone condition. 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% on the train set, of 59.3% on the development set, and of 58.0% on the test set. These high WER are mainly due to speech disfluencies and to adverse acoustic environments for some dialogues when, for example, users are calling from noisy streets with mobile phones. A "stop list" of 126 words[2] was used to remove unnecessary words which results in a WER of 33.8% on the train, of 45.2% on the development, and of 49.5% on the test.

Experiments are conducted with the two unsupervised classification methods (SVM / Gaussian) on the manual (TRS) and the automatic transcriptions only (ASR). We also propose to study the combination of both (TRS+ASR) in order to see if ASR errors can be supplied by the correct reference words.

The train set is used to compose a subset of discriminative words to elaborate a semantic space for each conversation of the test corpus with the basic TF-IDF-Gini method. In the experiments, the number of discriminative words has been varied from 800 to the total number of words contained in the train corpus (7,920 words). The test corpus contains 3,806 words (70.8% occur in the train corpus).

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

19 topic spaces with a different topic number ($\{5, \ldots, 300\}$) are also elaborated on the train corpus by using a LDA model made with the LDA Mallet Java implementation [30].

Then, for both configurations (semantic or topic vectors), a SVM classifier is learned with the LIBSVM library [31]. SVM parameters are optimized by cross validation on the train corpus. Finally, the SVM classifier is compared with a Gaussian classifier is trained based on the Bayes decision rule.

6. EXPERIMENTS AND RESULTS

In this section, the results with two different dialogue representations and two different classification methods are shown. Then, the representation space is reduced and the impact of this space reduction in terms of classification accuracy is discussed in section 6.3.

6.1. Impact of Dialogue representations

Figures 1-(a) and 1-(b) present the theme classification accuracies obtained by the TF-IDF-Gini and the LDA approaches on the test corpus for all transcription configurations (TRS/ASR) when varying the word extraction conditions (number of discriminative words and number of topics). We can see that the LDA-based method outperforms the best results obtained by the TF-IDF-Gini approach (see table 1).

Table 1: Theme classification accuracy on two dialogue representations with a SVM classifier (*Confidence of* $\pm 3.69\%$ *for LDA*)

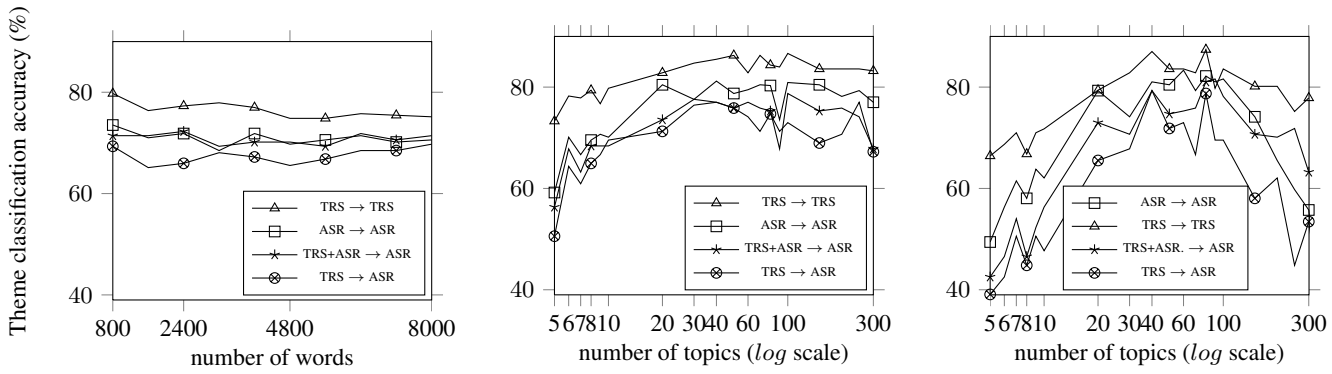
DATA		*BEST ACCURACY (%)*			
Train	**Test**	*#words*	**TF-IDF-Gini**	*#topics*	**LDA**
TRS	TRS	*800*	79.7	*100*	**86.6**
TRS	ASR	*8000*	69.7	*40*	**77.0**
ASR	ASR	*800*	73.5	*60*	**81.4**
ASR+TRS	ASR	*2400*	72.2	*100*	**78.7**

As expected, the best classification results are obtained by the TRS train / TRS test configuration (TRS $\rightarrow$ TRS) with a gain of 6.9 points with the LDA method. We can also note that the ASR test reached the best performance using the ASR training data condition. A gain of 7.9 points is noted with the LDA method compared to the TF-IDF-Gini approach on the automatic transcriptions. It seems clear that using comparable training and testing configurations allows to achieve the best classification performance, whether it be on manual or on automatic transcriptions.

We can finally note that the LDA approach performance has a tendency to fluctuate when varying the number of topics. This could be explained by the high Word Error Rate (WER) of the targeted corpus: indeed, the words chosen as *discriminative* in particular topic number conditions could be wrongly transcribed in a high proportion. This assumption is supported by analyzing the 90 topics condition (see figure 1). An important performance drop is observed for the ASR training conditions while a smaller performance lost is seen when using the reference transcriptions (TRS).

6.2. Impact of Classification methods

Figures 1-(b) and (c) show the theme classification accuracies obtained by the SVM and the Gaussian approaches on the test corpus for all transcription configurations (TRS/ASR). We can see that the Gaussian method outperforms the results obtained by the SVM approach no matter the condition studied (see table 2). As already seen in dialogue representation, the TRS $\rightarrow$ TRS configuration achieves the best results with a gain of 0.8 point using the Gaussian classifier



(a) **Term-frequency representation + SVM** (b) **Topic-based representation + SVM** (c) **Topic-based representation + Gaussian**

Fig. 1: Theme classification performance using 3 configurations: TF-IDF (a) and LDA (b) + SVM, and LDA (c) + Gaussian classifier.

method. If focusing on the ASR → ASR experience, the Gaussian method obtains a gain of 1.9 points in comparison to the SVM approach. We can finally notice that above 100 topics, the accuracy of the Gaussian classifier decreases. It could be explained by a lack of training data to correctly estimate a topic space of such size.

Table 2: Theme classification accuracy using the SVM and the Gaussian approaches (*Confidence of* $\pm 3.69\%$ *for SVM*)

DATA		*BEST ACC. (%) SVM*		*BEST ACC. (%) Gaussian*			
Train	**Test**	*#topics*	**Test**	*#topics*	**Dev**	*#topics*	**Test**
TRS	TRS	*100*	86.6	*80*	92.2	*80*	**87.4**
TRS	ASR	*40*	77.0	*40*	84.5	*80*	**79.3**
ASR	ASR	*60*	81.4	*60*	86.6	*80*	**83.3**
ASR+TRS	ASR	*100*	78.7	*90*	89.7	*80*	**81.6**

6.3. Impact of the space reduction

Figures 2-(a) and 2-(b) present the theme classification performance obtained on the manual transcription condition (TRS → TRS) using the Gaussian-based Bayes classification approach.

The original precision curve (dashed line) represents the results obtained with the LDA-based method using the original topic space, already presented in figure 1-(b) (TRS → TRS). These results are compared with those obtained with a topic space of reduced size (triangle dots). These reductions are performed from a Principal Component Analysis (PCA) on topic spaces that have a size greater than n ($n = 40$ in figure 2-(a) or $n = 80$ in figure 2-(b)). The last line (square dots) represents the results obtained with the original topic space of exact size n.

Let's consider the topic space size of 80 dimensions in figure 2-(a) (40 dimensions). We can see that the original classification accuracy is about 87%. Then, when focusing on the PCA reduction (LDA+PCA → space size=40), the accuracy reaches about 85%. This precision has been obtained by reducing the number of topics from 80 to 40 dimensions. Thus, we can notice that the precision decreases using this new reduced space representation. For all the other studied dimension sizes ($n \neq 80$), the reduction of the topic space sizes to n improves the results for both cases ($n = 40$ in or $n = 80$). Nonetheless, we have to note that the classification accuracy never reaches the best results obtained with the original LDA topic space size (square dots).

We can conclude that the reduction permits to cluster topics and to improve the results. The fact that we artificially increase the number of topics (granularity) without increasing the number of conversations in the training corpus for each theme, leads to decrease the variability within themes but does not allow to reach the accuracy of the optimal exact topic space size (n).

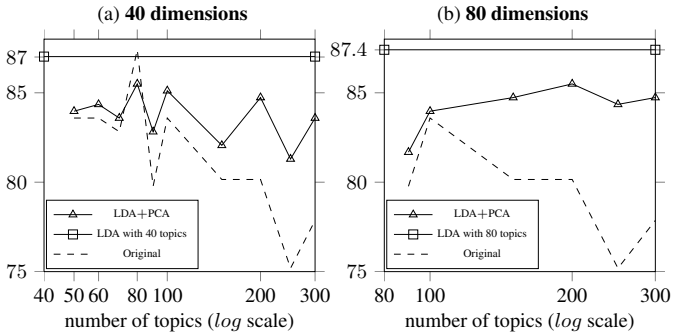


Fig. 2: Theme classification performance (accuracy %) with models m ($|m| > n$) for $n = 40$ (a) and $n = 80$ (b).

7. CONCLUSIONS

In this paper, we presented an architecture to identify conversation themes using two different dialogue representations and classification methods. We showed that the proposed topic representation using a LDA-based method outperforms the classification results obtained by the classical TF-IDF-Gini approach. The classification accuracy reaches 86.6% on manual transcriptions and 81.4% on automatic transcriptions with a respective gain of 6.9 and 7.9 points.

The second part of the work focused on choosing the best classification method. We highlighted that the intuitions about the Gaussianity of the theme classes and the equality of the class covariances discussed in this paper are effective. Thus, the topic representation using a Gaussian classifier method outperforms the classification results obtained by the classical SVM approach. The accuracy reaches 87.4% on manual transcriptions and 84.4% on highly imperfect automatic transcriptions with a respective gain of 0.8 and 1.9 points.

In the last part of this study, we demonstrated that the space reduction (LDA+PCA) improves the results obtained with the dialogues in the original topic space. Nonetheless, we also pointed out that this reduction does not allow to achieve the original results obtained with the optimal exact topic space size.

8. 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] 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.

[3] Stephen Robertson, "Understanding inverse document frequency: on theoretical arguments for IDF," *Journal of Documentation*, vol. 60, no. 5, pp. 503–520, 2004.

[4] Tao Dong, Wenqian Shang, and Haibin Zhu, "An improved algorithm of bayesian text categorization," *Journal of Software*, vol. 6, no. 9, pp. 1837–1843, 2011.

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

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

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

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

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

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

[11] Bernhard E. Boser, Isabelle M. Guyon, and Vladimir N. Vapnik, "A training algorithm for optimal margin classifiers," in *5th annual workshop on Computational learning theory*, 1992, pp. 144–152.

[12] Klaus-Robert Müller, Alexander Smola, Gunnar Rätsch, Bernhard Schölkopf, Jens Kohlmorgen, and Vladimir Vapnik, "Predicting time series with support vector machines," *ICANN'97*, pp. 999–1004, 1997.

[13] Thorsten Joachims, "Transductive inference for text classification using support vector machines," in *Machine learning-international workshop then conference*. Morgan Kaufmann Publishers, Inc., 1999, pp. 200–209.

[14] Jian hua Yeh and Chun hsing Chen, "Protein remote homology detection based on latent topic vector model," in *International Conference on Networking and Information Technology (ICNIT)*, 2010, pp. 456–460.

[15] Mounir Zrigui, Rami Ayadi, Mourad Mars, and Mohsen Maraoui, "Arabic text classification framework based on latent dirichlet allocation.," *CIT*, vol. 20, no. 2, pp. 125–140, 2012.

[16] Samuel Kim, Shrikanth Narayanan, and Shiva Sundaram, "Acoustic topic model for audio information retrieval," in *Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2009, pp. 37–40.

[17] Mohamed Morchid, Richard Dufour, and Linars Georges, "Event detection from image hosting services by slightly-supervised multi-span context models," in *CBMI*. 2013, IEEE.

[18] Sheng Tang, Jintao Li, Yongdong Zhang, Cheng Xie, Ming Li, Yizhi Liu, Xiufeng Hua, Yan-Tao Zheng, Jinhui Tang, and Tat-Seng Chua, "Pornprobe: an lda-svm based pornography detection system," in *International Conference on Multimedia*, 2009, pp. 1003–1004.

[19] J. I. Sheeba and K. Vivekanandan, "Improved keyword and keyphrase extraction from meeting transcripts," *International Journal of Computer Applications*, vol. 52, no. 13, pp. 11–15, 2012.

[20] Najim Dehak, Patrick J Kenny, Réda Dehak, Pierre Dumouchel, and Pierre Ouellet, "Front-end factor analysis for speaker verification," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 4, pp. 788–798, 2011.

[21] Pierre-Michel Bousquet, Driss Matrouf, and Jean-François Bonastre, "Intersession compensation and scoring methods in the i-vectors space for speaker recognition.," in *INTERSPEECH*, 2011, pp. 485–488.

[22] Eric P Xing, Michael I Jordan, Stuart Russell, and Andrew Ng, "Distance metric learning with application to clustering with side-information," in *Advances in neural information processing systems*, 2002, pp. 505–512.

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

[24] 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.

[25] 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.

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

[27] Sergios Petridis and Stavros J Perantonis, "On the relation between discriminant analysis and mutual information for supervised linear feature extraction," *Pattern Recognition*, vol. 37, no. 5, pp. 857–874, 2004.

[28] 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.

[29] 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.

[30] Andrew K. McCallum, "Mallet: A machine learning for language toolkit," 2002.

[31] Chih-Chung Chang and Chih-Jen Lin, "Libsvm: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, pp. 27, 2011.