Theme Identification in Human-Human Conversations with Features from Specific Speaker Type Hidden Spaces

Mohamed Morchid[†], Richard Dufour[†], Mohamed Bouallegue[†],
Georges Linarès[†] and Renato De Mori[†‡]

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

Abstract

This paper describes a research on topic identification in a real-world customer service telephone conversations between an agent and a customer. Separate hidden spaces are considered for agents, customers and the combination of them. The purpose is to separate semantic constituents from the speaker types and their possible relations. Probabilities of hidden topic features are then used by separate Gaussian classifiers to compute theme probabilities for each speaker type. A simple strategy, that does not require any additional parameter estimation, is introduced to classify themes with confidence indicators for each theme hypothesis. Experimental results on a real-life application show that the use of features from speaker type specific hidden spaces capture useful semantic contents with significantly superior performance with respect to independent word-based features or a single set of features. Experimental results also show that the proposed strategy makes it possible to perform surveys on collections of conversations by automatically selecting processed samples with high theme identification accuracy.

Index Terms: Spoken language understanding, human/human telephone conversation analysis, LDA, topic identification

1. Introduction

The automatic analysis of human/human telephone conversations has received growing attention suggesting new interesting scientific challenges and possible new call center applications such as customer care services (CCS). An important aspect of the problem is the automatic detection of themes that are considered relevant for a specific application domain. The term *theme* is used in this paper to indicate conversation topics in the application domain as opposed to hidden topic spaces where hidden features are computed. Various features and approaches reviewed, in [1], have been proposed in the literature for theme hypothesization in spoken and text documents. The approaches use supervised and unsupervised feature selection methods. An interesting possibility, among the unsupervised approaches, is to use hidden features obtained with Latent Dirichlet Allocation (LDA). These features capture dependencies among conversation words that may describe useful semantic contents.

In these conversations, the agent attempts to follow a protocol defined in the application documentation, while the customer may have an unpredictable behavior mixing domain relevant with domain irrelevant information with repetitions, self-

corrections and other disfluencies. Furthermore, the customer may speak in a very noisy acoustic environment such as a subway station. As a result, the ASR hypotheses generated by an automatic speech recognition (ASR) system may be affected by high word error rates (WER). This motivates the use of only 1-best ASR hypotheses considering that hidden topics generated with hypothesized words still maintain a sufficient discriminative power for topic hypothesization. For the real-life application considered in this paper, in spite of high ASR word error rates (WER), high accuracy in theme hypothesization has been recently reported in [2]. These results have been obtained with a probabilistic classifier using features obtained with a hidden LDA space common to agents and speakers. Given the evident difference between speaker types and acoustic environments, the possibility of characterizing more semantically relevant dependencies by using more hidden topic spaces is investigated in this paper. Specific LDA hidden spaces are considered respectively for the agent, the speaker, and the combination of the two. In addition to that, as different noise environments may result in large variations of the WER of each conversation, the possibility is considered to take into account the WER of each conversation of the train set for building a specific hidden space.

The novel contributions of this paper are summarized in the following. Theme classification results are obtained with three different hidden spaces and specific space sizes using a Gaussian-based Bayesian classifier. Performance for each hidden space, different train conditions and different numbers of hidden topics per space have been compared.

The possibility of exploiting partial or total coherence of results obtained with the different hidden spaces is considered. As part of the research requires investigating the effect of varying the number of dimensions in each of the considered hidden spaces, the results reported in this paper refer only to the use of the 1-best ASR word hypotheses. The possibility of using lattices of word hypotheses will be considered in future work.

Having a separate classification process for each hidden space feature set makes it possible to compare multiple view classification results and use their partial or total consensus as a confidence indicator of theme hypotheses. A considerable proportion of high accuracy classifications is observed for examples with a large consensus. For these examples, automatic classification can be accepted without any further supervision. A significant improvement with respect to the results reported in [2] using a single hidden space is observed with a simple strategy suggested by the results of the development set and consisting in selecting the theme receiving the maximum consensus in the three spaces and the theme hypothesized in the

agent hidden space in case of total disagreement.

For the small proportion of ambiguously classified conversations simple confidence indicators based on consensus are derived for signaling the possibility of validation by a human expert to obtain a global reliable classification with a minor effort.

The paper is organized as follows. Section 2 presents the related work. The features used for theme hypothesization are described in section 3. Section 4 introduces the classification methods. Section 5 reports experimental results before concluding in section 6.

2. Related work

A recent review on topic identification in spoken documents can be found in [1]. The review discusses, among other things, methods such as Latent Semantic Analysis (LSA) [3, 4], Probabilistic LSA (PLSA) [5], and (LDA) [6], for building a higher-level representation of documents in a topic space. A document is then represented by a bag-of-words ignoring any word order. These methods demonstrated good performance in tasks such as sentence [7] or keyword [8] extraction.

In particular, LDA has recently been applied in various domains, such as biology [9], text classification [10], stylometry [11], audio information retrieval [12], social event detection [13] or image processing [14]. More recently, LDA hidden topic features have been combined with a support vector machine (SVM) classifier to assign spoken documents to broad classes [15]. In this work, it was observed that using the LDA topic weights as features obtained with a lattice of word hypotheses obtained by an ASR system, required additional features to outperform a baseline system using the other features.

Gaussian classifiers with the Mahalanobis metric distance [16] have been applied for speaker recognition tasks [17] and audio recognition of speaker identities [18].

3. Features used for theme hypothetization

A dialogue d in a corpus D is described by features belonging to a vocabulary $\mathbf{V} = \{w_1, \ldots, w_N\}$ of size N. For the comparative experiments described in this paper, a set of discriminative words [19] have been selected. In order to define elements of a document feature vector, a discriminative term δ is defined for a word w in a theme t as $\delta_t^w = tf_t(w) \times idf(w) \times gini(w)$ [20].

$$gini(w) = 1 - \sqrt{\sum_{i=1}^{|\mathbf{T}|} p_i^2} \ ,$$

where p_i is the probability that a word w is generated by the ith theme and $\mathbf{T}$ (where $t \in \mathbf{T}$) is the the set of all themes.

Then the words having the highest scores Δ for all the themes $\mathbf{T}$ are extracted and constitute a discriminative word subset $\mathbf{V_\Delta}$ (each theme $t \in \mathbf{T}$ has its own score δ_t) and its own frequency γ in the model f:

$$\gamma_f^t = \frac{\#d \in t}{\#d \in D} \ .$$

Note that the same word w can be present in different themes, but with different scores (TF-IDF-Gini) depending of its relevance in the theme. The just introduced features are defined in a space in which a dimension corresponds to a word. In the next subsection, a different feature is introduced. These features are defined in a hidden space of hidden topics. Notice that these features are used only in a classifier introduced for comparison.

3.1. Hidden topic features

Hidden spaces are obtained using the unigram probabilities of each word w in the entire application vocabulary $\mathbf{V}$. For every spoken document d of a corpus D, a parameter θ is drawn from a Dirichlet distribution with parameter α in the LDA based model introduced in [6].

The Gibbs sampling algorithm [21] is used to estimate features of a dialogue d in each considered LDA hidden space. This algorithm is based on the Markov Chain Monte Carlo (MCMC) method and allows us to obtain samples of the distribution parameters θ knowing a word w of a document and a hidden topic z.

A hidden space of n topics is obtained with, for each theme z, the probability of each word w of $\mathbf{V}$ knowing z ($P(w|z) = V_z^w$) (see figure 1).

A feature vector V_d^z is then obtained in each hidden space. The kth feature $V_d^z[k]$ (where $1 \leq k \leq n$) in the vector is the probability of hidden topic z_k given a conversation d:

$$V_d^z[k] = P(z_k|d).$$

Hidden topics are obtained with the approach proposed in this paper using separately the entire conversation, the turns of the agent, and the turns of the customer.
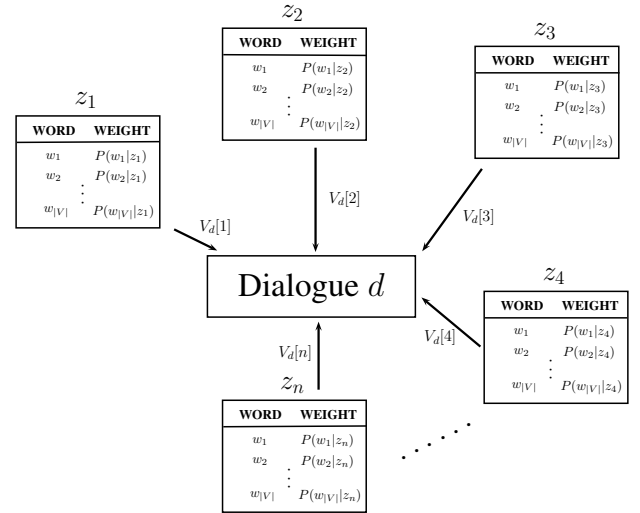

Figure 1: Mapping of a dialogue in the topic space.

4. Classification methods

Hidden topic probabilities computed for a document are used as features for theme hypothesization. Two classifiers are considered for this purpose in order to compare the results they provide. The first classifier is a Gaussian classifier while the second is an SVM.

4.1. Gaussian based Bayes classifier

The homoscedastic Gaussian Bayesian classifier [22] is based on two simple assumptions, namely that distributions of theme classes are Gaussian and the covariances of these classes are equal. The Gaussian classifier assigns a theme label to dialogue d with a Bayesian decision rule with a scoring metric. Given a training dataset of dialogues D, let $\mathbf{W}$ denote the within dia-

logue covariance matrix defined by:

$$\mathbf{W} = \sum_{k=1}^{K} \frac{n_k}{n} \mathbf{W_k} = \frac{1}{n} \sum_{k=1}^{K} \sum_{i=0}^{n_k} \left(x_k^i - \overline{x_k} \right) \left(x_k^i - \overline{x_k} \right)^T \tag{1}$$

where $\mathbf{W_k}$ is the covariance matrix of the kth theme C_k, n_k is the number of utterances for the theme k, n is the total number of dialogues in the training dataset, x_i^k are the training dialogues of theme k, $\overline{x_k}$ is the mean of all dialogues of the kth theme and K is the number of themes. Dialogues do not provide the same contribution to the covariance matrix. For this reason, the term $\frac{n_k}{n}$ is introduced in equation 1.

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

$$
\begin{aligned}
k_{\text{Bayes}} &= \arg \max_k \; \mathcal{N} \left(x \mid \overline{x_k}, \mathbf{W} \right) \\
&= \arg \max_k \left\{ -\frac{1}{2} \left(x - \overline{x_k} \right)^T \mathbf{W}^{-1} \left(x - \overline{x_k} \right) + a_k \right\}
\end{aligned}
\tag{2}
$$

where $\overline{x_k}$ is the centroid (mean) of theme k, $\mathbf{W}$ is the within theme covariance matrix defined in equation 1, $\mathcal{N}$ denotes the normal distribution and a_k is the log prior probability of the theme membership defined as $a_k = \log\left(P(C_k)\right)$. It is worth noting that, with these assumptions, the Bayesian approach is similar to the Fisher's geometric approach since x is assigned to the nearest centroid's class, according to the Mahalanobis [16] metric of $\mathbf{W}^{-1}$ computed as follows:

$$k_{\text{Bayes}} = \arg \max_k \left\{ -\frac{1}{2} ||x - \overline{x_k}||^2_{\mathbf{W}^{-1}} + a_k \right\} \tag{3}$$

4.2. SVM classification

SVM classifiers are used for the purpose of comparison. They map hidden topic features into a space of higher dimension and make decisions in this new space. As theme classification requires a multi-class classifier, the SVM *one-against-one* method is chosen with a linear kernel. This method gives a better testing accuracy than the *one-against-rest* method [23]. 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 themes. As a result, a set of $T(T-1)/2$ binary classifiers are trained and used for testing. The binary classifier $C_{i,j}$ is trained with data labelled with t_i for a positive class and t_j for negative one $(i \neq j)$.

Given a dialogue d in the test corpus, if $C_{i,j}$ classifies d in theme t_i, then a voting score for the class t_i is incremented by one. Otherwise, the score for the theme t_j is increased by one. Eventually, the dialogue d is assigned to the theme that received the highest score.

5. Experiments

5.1. Experimental protocol

The corpus of the DECODA project [24] has been used for the theme identification experiments described in this section.

This corpus is composed of 1,067 telephone conversations from the call centre of the public transportation service in Paris.

The corpus is split into a train set (740 dialogues) and a test set (327 dialogues). Conversations have been manually transcribed and labeled with one theme label corresponding to the principal concern mentioned by the customer. The semantic annotation consists in 8 conversation themes: *problems of itinerary*, *lost and found*, *time schedules*, *transportation cards*, *state of the traffic*, *fares*, *infractions* and *special offers*. A portion of the train set (175 dialogues) is also used as a development set for selecting the dimension of the hidden topic spaces. All hidden spaces were obtained with the manual transcriptions of the train set. The number of turns in a conversation and the number of words in a turn are highly variable. The majority of the conversations have more than ten turns. The turns of the customer tend to be longer (> 20 words) than those of the agent and are more likely to contain out of vocabulary words that are often irrelevant for the task.

The ASR system used for the experiments is the LIA-Speeral system [25] with 230,000 Gaussians in the triphone acoustic models. Model parameters were estimated with maximum *a posteriori* probability (MAP) adaptation from 150 hours of another corpus of telephone speech. The vocabulary contains 5,782 words. A 3-gram language model (LM) was obtained by adapting with the manual transcriptions of the train set a basic LM. An initial set of experiments was performed with this system resulting in an overall WER on the train set of 45.8% and on the test set of 58.0%. These high WER are mainly due to speech disfluencies and to adverse acoustic environments for some dialogues when, for example, users are calling from train stations or noisy streets with mobile phones. Furthermore, the signal of some sentences is clipped or with low signal to noise ratio. A "stop list" of 126 words[1] was used to remove unnecessary words resulting in a WER of 33.8% on the train set and of 49.5% on the test set.

Experiments are conducted using train data represented by the manual transcriptions only (TRS), the automatic transcriptions only (ASR), and by dividing the word probabilities of a conversation with the WER of the entire conversation (A+WER). The conditions indicated by the abbreviations between parentheses are considered for the development (Dev) and the test (Test) sets.

Classifications are considered in the above mentioned train and test conditions using features from hidden spaces of respectively agent only (AGENT), customer only (CUSTOMER) and the combination of the two (AG-CUST).

For the sake of comparison, theme hypothesization was performed with the two unsupervised classification methods, namely SVM and Gaussian, using hidden topic features estimated with word probabilities. For the sake of comparison, TF-IDF-Gini features were also evaluated with an SVM classifier.

As the agents annotate conversations based on what they considered as the most important theme if multiple themes are mentioned, the corpus annotation used in the experiments is based on the agent annotations with minor corrections made only when unquestionable agent errors, mostly caused by stress, are observed.

It was observed that the train corpus vocabulary contains 7,920 words while the test corpus contains 3,806 words, only 70.8% of them occur in the train corpus. A subset of the 800 most discriminative words according to the Gini score was extracted to compose 800 TF-IDF-Gini features used with an SVM classifier.

For each train condition and feature type, a set of 19 hidden

---

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

topic spaces with a different topic number ({5, 6, 7, ..., 10, 20, ..., 100, 150, ..., 300}) was built using the train corpus. For the data in the test corpus, a feature vector is computed by mapping each dialogue with each topic space. The topic spaces are obtained with the LDA Mallet Java implementation[2].

SVM classifiers are trained with the LIBSVM library [26]. SVM parameters are optimized by cross validation on the train corpus. The parameters of the Gaussian classifier are estimated using the Bayes decision rule.

5.2. Results

The separate results of the classification experiments with LDA hidden topic features and Gaussian classifier for each speaker type are summarized in Table 1. Each row corresponds to the type of data transcriptions in the train and the development or the test sets. Each column corresponds to the speaker type (AGENT, CUSTOMER, AG-CUST) and the set (Dev, Test) on which the evaluations are performed. The numbers reported on columns labelled with $|t|$ indicate the sizes of the hidden topic spaces corresponding to the middle of the interval of sizes in which the classification performance exhibits small variations for the Dev corpus. A very simple consensus strategy was then applied. It consists in selecting the theme that receives the highest classification score in at least two hidden spaces and selecting the hypotheses generated using only features from the AG-CUST space when there is no consensus among the three classifiers. The confidence interval for the test set is $\pm$ 3.69%.

As expected, the results show that the best results using the ASR transcriptions of the development set are obtained using ASR transcriptions of the train set. Results also show that there is no advantage in taking into account the WER evaluated on each conversation. Moreover, the accuracy observed for the agent conversation turns is higher than that of the customer and the combination of both speakers. This may be due to difference in the environments and speaking styles. The influence of these differences is somehow alleviated by the fact that features are computed in specific hidden spaces. An analysis of the train set conversations shows that the agents tend to use similar expressions for certain explanations because they have been trained to follow a protocol. Nonetheless, customers often repeat details about their problem providing information that is complementary with respect to agent answers.

Table 1: Comparative theme hypothesization accuracy results using different hidden topic spaces for different train situation and different types of speakers.

DATA	CUSTOMER			AGENT			AG-CUST		
train/test	$\|t\|$	Dev	Test	$\|t\|$	Dev	Test	$\|t\|$	Dev	Test
TRS/TRS	80	89.7	**84.7**	80	91	**85.3**	80	92.5	**86.8**
ASR/ASR	80	85.4	**79.4**	70	87.4	**80.4**	80	84.8	**82.5**
A.+WER/A.	80	76	**72.3**	60	78.1	**73.1**	100	75.6	**75.2**
TRS/ASR	100	82.4	**74.4**	100	85.4	**77.6**	80	84.8	**78.3**

Table 2 shows the results obtained using ASR transcriptions and a Gaussian classifier compared with the use of the same features with an SVM classifier and with the results obtained with TF-IDF-Gini discriminative features with an SVM classifier. The theme hypothesization strategy denoted MAJ1 consists in selecting the theme receiving the maximum consensus in the three spaces and the theme hypothesized in the agent hidden space in case of total disagreement. The results obtained with the proposed approach are superior to those reported in [2, 27].

Furthermore, the improvements of the proposed approach with respect to the competitors are now statistically significant.

Table 2: Comparison between accuracies of different feature types and different classifiers with hidden topic features.

DATA	TF-IDF-GINI	SVM	AG-CUST	MAJ1
train/test	**Test**	**Test**	**Test**	**Test**
TRS/TRS	74.1	85.5	86.8	**87.2**
ASR/ASR	64.5	80.4	82.5	**84.1**
TRS/ASR	58.4	72	78.3	**78.5**

A strategy based on classifier consensus in the ASR/ASR conditions has been used to compose the following test sets. MAJ3 is the subset of the test made of the conversations for which the same theme has been hypothesized for CUSTOMER, AGENT, and the combination of the two. MAJ2 is the subset of the test set containing all the conversations for which with at least two classifiers have hypothesized the same theme. If the hypotheses generated by the three classifiers are all different, then the decision made in the AG-CUST hidden space is pooled with the decisions made in the two previous cases to form the set MAJ1. Table 3 shows, for each consensus type, the theme hypothesization accuracy and the coverage measured by the proportion of conversations considered in the corresponding set. The results show that the consensus strategy provides very useful confidence indicators without any specific strategy training.

Table 3: Theme hypothesization accuracy and coverage for different strategy situations with the best train condition (ASR→ASR) (topic space size=80).

DATA	MAJ3		MAJ2		MAJ1	
train/test	Acc.	Cov.	Acc.	Cov.	Acc.	Cov.
ASR/ASR	95.1	57.6	86.3	92.2	**84.1**	100

6. Discussions and Conclusions

The results reported in Table 2 show that the use of features from speaker type specific hidden LDA spaces capture useful semantic contents with significantly superior performance than independent TF-IDF features. Speaker type specific LDA features sets provide better results than a single set of features in a common LDA space using a Gaussian Bayesian classifier.

The results reported in Table 3 show that useful theme classification confidence indicators can be conceived and used in simple strategy that does not require any parameter estimation. With this consensus based strategy it is possible to perform a survey on a collection of conversations by selecting automatically processed samples with large consensus. In this way, in spite of very high WER, it is possible to compose a survey with samples annotated with 95% accuracy and covering more than 57% of the entire population. Such a high accuracy with good coverage is a great advantage with respect to previous approaches applied to the same corpus [28]. With such a high accuracy it will be possible to estimate proportions of user problems in specific time intervals or traffic situations. If these proportions are estimated on a sufficiently extended survey, user problems and concerns can be monitored to make suitable decisions for improving the service.

Future work will focus on the detection of multiple themes in a conversation, detection of out of domain conversation detection, hypothesization of specific mentions of entities, telephone numbers, call transfers, service names, user satisfaction expression for which methods based on conversation segmentation, such as the one described in [19], are more suitable.

7. References

[1] T. J. Hazen, *Topic identification ch. 12 of G. Tur and R. De Mori Eds. Spoken Language Understanding*. John Wiley & Sons, 2011.

[2] M. Morchid, R. Dufour, P.-M. Bousquet, M. Bouallegue, G. Linarès, and R. De Mori, "Improving dialogue classification using a topic space representation and a gaussian classifier based on the decision rule." ICASSP'14, 2014.

[3] S. Deerwester, S. Dumais, G. Furnas, T. 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.

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

[5] T. Hofmann, "Probabilistic latent semantic analysis," in *Proc. of Uncertainty in Artificial Intelligence, UAI ' 99*. Citeseer, 1999, p. 21.

[6] D. Blei, A. Ng, and M. Jordan, "Latent dirichlet allocation," *The Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

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

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

[9] J. hua Yeh and C. 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.

[10] D. M. Blei and J. D. McAuliffe, "Supervised topic models," *arXiv preprint arXiv:1003.0783*, 2010.

[11] Arun, Saradha, V. Suresh, Murty, and C. E. Veni Madhavan, "Stopwords and Stylometry : A Latent Dirichlet Allocation Approach," in *NIPS Workshop on Applications for Topic Models: Text and Beyond*, Whistler, Canada, 2009.

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

[13] M. Morchid, R. Dufour, and G. Linarès, "Event detection from image hosting services by slightly-supervised multi-span context models," in *International Workshop on Content-Based Multimedia Indexing (CBMI'13)*, 2013.

[14] S. Tang, J. Li, Y. Zhang, C. Xie, M. Li, Y. Liu, X. Hua, Y.-T. Zheng, J. Tang, and T.-S. Chua, "Pornprobe: an lda-svm based pornography detection system," in *International Conference on Multimedia*, 2009, pp. 1003–1004.

[15] J. Wintrode, "Using latent topic features to improve binary classification of spoken documents," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. IEEE, 2011, pp. 5544–5547.

[16] E. P. Xing, M. I. Jordan, S. Russell, and A. Ng, "Distance metric learning with application to clustering with side-information," in *Advances in neural information processing systems*, 2002, pp. 505–512.

[17] P.-M. Bousquet, D. Matrouf, and J.-F. Bonastre, "Inter-session compensation and scoring methods in the i-vectors space for speaker recognition." in *INTERSPEECH*, 2011, pp. 485–488.

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

[19] M. Morchid, G. Linarès, M. El-Beze, and R. De Mori, "Theme identification in telephone service conversations using quaternions of speech features," in *Conference of the International Speech Communication Association (INTERSPEECH), Lyon, France*. ISCA, 2013.

[20] T. Dong, W. Shang, and H. Zhu, "An improved algorithm of bayesian text categorization," *Journal of Software*, vol. 6, no. 9, pp. 1837–1843, 2011.

[21] T. Griffiths and M. Steyvers, "A probabilistic approach to semantic representation," in *24th annual conference of the cognitive science society*. Citeseer, 2002, pp. 381–386.

[22] S. Petridis and S. 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.

[23] G.-X. Yuan, C.-H. Ho, and C.-J. Lin, "Recent advances of large-scale linear classification," vol. 100, no. 9, pp. 2584–2603, 2012.

[24] F. Bechet, B. Maza, N. Bigouroux, T. Bazillon, M. El-Beze, R. De Mori, and E. Arbillot, "Decoda: a call-centre human-human spoken conversation corpus." LREC'12, 2012.

[25] G. Linarès, P. Nocéra, D. Massonie, and D. Matrouf, "The lia speech recognition system: from 10xrt to 1xrt," in *Text, Speech and Dialogue*. Springer, 2007, pp. 302–308.

[26] C.-C. Chang and C.-J. Lin, "Libsvm: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.

[27] M. Morchid, R. Dufour, and G. Linarès, "A LDA-based topic classification approach from highly imperfect automatic transcriptions," in *LREC'14*, 2014.

[28] B. Maza, M. El-Beze, G. Linares, and R. De Mori, "On the use of linguistic features in an automatic system for speech analytics of telephone conversations," in *Interspeech'11*, 2011.