

Un Sous-espace Thématique Latent pour la Compréhension du Langage Parlé

Mohamed Bouaziz^{1,2} Mohamed Morchid¹ Pierre-Michel Bousquet¹
Richard Dufour¹ Killian Janod¹ Waad Ben Kheder¹ Georges Linarès¹

(1) LIA, 339 Chemin des Meinajaries, 84140 Avignon, France

(2) EDD, 28 Boulevard de Port-Royal, 75005 Paris, France

mohamed.bouaziz@alumni.univ-avignon.fr, {prénom.nom}@univ-avignon.fr

RÉSUMÉ

Les applications de compréhension du langage parlé sont moins performantes si les documents transcrits automatiquement contiennent un taux d'erreur-mot élevé. Des solutions récentes proposent de projeter ces transcriptions dans un espace de thèmes, comme par exemple l'allocation latente de Dirichlet (LDA), la LDA supervisée ainsi que le modèle *author-topic* (AT). Une représentation compacte originale, appelée *c*-vector, a été récemment introduite afin de surmonter la difficulté liée au choix de la taille de ces espaces thématiques. Cette représentation améliore la robustesse aux erreurs de transcription, en compactant les différentes représentations LDA d'un document parlé dans un espace réduit. Le défaut majeur de cette méthode est le nombre élevé de sous-tâches nécessaires à la construction de l'espace *c*-vector. Cet article propose de corriger ce défaut en utilisant un cadre original fondé sur un espace de caractéristiques robustes de faible dimension provenant d'un ensemble de modèles AT considérant à la fois le contenu du dialogue parlé (les mots) et la classe du document. Les expérimentations, conduites sur le corpus DECODA, montrent que la représentation proposée permet un gain de plus de 2.5 points en termes de conversations correctement classifiées.

ABSTRACT

A Latent Topic-based Subspace for Spoken Language Understanding.

Performance of spoken language understanding applications declines when spoken documents automatically transcribed contain high Word Error Rates (WER). Recent solutions propose to map these automatic transcriptions in topic-based representations, such as Latent Dirichlet Allocation (LDA), supervised LDA and author-topic (AT) models. An original compact representation, called *c*-vector, has recently been introduced to walk around the tricky choice of the number of latent topics in these topic-based representations. This representation increases the robustness of document classification with respect to transcription errors by compacting different LDA representations of a same speech document in a reduced space and then compensate most of the noise of the document representation. The main drawback of this method is the number of sub-tasks needed to build the *c*-vector space. This paper proposes to tackle this drawback using an original framework in a robust low dimensional space of features from a set of AT models, considering not only the dialogue content (words), but also the class related to the document. Experiments, conducted on the DECODA corpus, show that the original LTS representation allows a gain of more than 2.5 points in terms of correctly labeled conversations.

MOTS-CLÉS : Modèle Author-Topic, Analyse factorielle, *c*-vecteur, Classification de documents.

KEYWORDS: Author-Topic model, Factor analysis, *c*-vector, Document classification.

1 Introduction

Les performances des applications de compréhension du langage diminuent lorsque les transcriptions automatiques des documents parlés contiennent un grand nombre d’erreurs. L’étape de reconnaissance de la parole doit faire face à des difficultés d’origines (conditions de prise du son, bruits environnementaux, nombreuses disfluences...).

Des travaux récents portant sur l’analyse des conversations parlées, l’analyse de la parole, l’identification du sujet et la segmentation, peuvent être trouvés dans (Eisenstein & Barzilay, 2008; Lagus & Kuusisto, 2002; Tur & De Mori, 2011; Hazen, 2011; Melamed & Gilbert, 2011) et (Purver, 2011). Une manière efficace pour améliorer la robustesse des Systèmes de Reconnaissance Automatique de la Parole (SRAP) est de projeter les conversations dans un espace abstrait permettant une classification efficace de ces dialogues dans un espace caché robuste. De nombreux espaces thématiques non supervisés ont été proposés afin de représenter efficacement le contenu d’un dialogue comme l’allocation latente de Dirichlet (*Latent Dirichlet Allocation* ou LDA) (Blei *et al.*, 2003) et les modèles *author-topic* (AT) (Rosen-Zvi *et al.*, 2004). (Morchid *et al.*, 2014a) et (Morchid *et al.*, 2014c) ont respectivement surmonté deux difficultés :

- réussir à choisir le nombre de thèmes optimal de l’espace thématique en utilisant plusieurs représentations latentes obtenues en variant la taille de l’espace LDA puis en compactant ces représentations à l’aide de l’analyse factorielle (Dehak *et al.*, 2011) (diverses sous-tâches sont requises),
- contraindre les méthodes de recherche de thèmes latents, qui sont non supervisées pour la plupart, à se rapprocher des thématiques recherchées dans la tâche. Dans ce but, le modèle *author-topic* a été défini pour prendre en considération à la fois le contenu des documents (*i.e.* les mots), la thématique du document, lorsque celle-ci est connue, mais aussi la distribution des mots sachant la thématique qui est modélisée par une relation latente.

Cet article propose tout d’abord de traiter conjointement les deux problèmes ci-dessus en apprenant un grand nombre de modèles AT et, ensuite, en extrayant de ces représentations des vecteurs compacts au moyen de l’analyse factorielle. Cette approche nécessite de nombreux pré-traitements ou projections (réseau de neurones profond (May *et al.*, 2015), UBM-GMM, normalisation...), les gains les plus importants étant observés sur des données très bruitées (Bousquet *et al.*, 2011). Néanmoins, notons que des représentations telles que les modèles AT ne contiennent qu’une faible portion de bruit due à la taille réduite de cette représentation comparativement à un espace LDA classique. Ainsi, ce papier propose de considérer les différents modèles AT comme un sous-espace de caractéristiques commun et de compacter ces différentes représentations afin d’extraire directement un vecteur robuste de faible dimension (super-vecteur).

La suite de ce papier est organisée comme suit. Les approches proposées sont décrites dans la section 2. La section 3 présente le protocole expérimental et expose les résultats. Enfin, la section 5 conclut ce travail et offre quelques perspectives.

2 Approche proposée

L’approche originale proposée, appelée “sous-espace thématique latent” (*Latent Topic-based Subspace* ou LTS), est comparée avec la représentation *c*-vecteur (Morchid *et al.*, 2014a). Les deux approches apprennent un ensemble d’espaces AT détaillés dans la section 2.1, puis projettent les différents

documents dans des espaces thématiques. Enfin, une compression de ces représentations est effectuée avec l'analyse factorielle pour la représentation fondée sur le c -vecteur, et avec la Décomposition en Éléments Propres (*Eigenvalues Decomposition* ou EVD) pour le LTS. La section 2.2 décrit l'approche c -vecteur illustrée dans la figure 1-(a)-(b)-(c), l'approche proposée (LTS) étant présentée dans la section 2.3. En ce qui concerne la technique LTS, les divers espaces thématiques sont considérés comme des sous-espaces latents homogènes, ce qui évite de projeter les documents dans un GMM. En outre, les super-vecteurs qui composent le LTS sont compressés à l'aide d'une EVD simple afin d'extraire une représentation robuste du document. Ces méthodes sont décrites dans la section suivante.

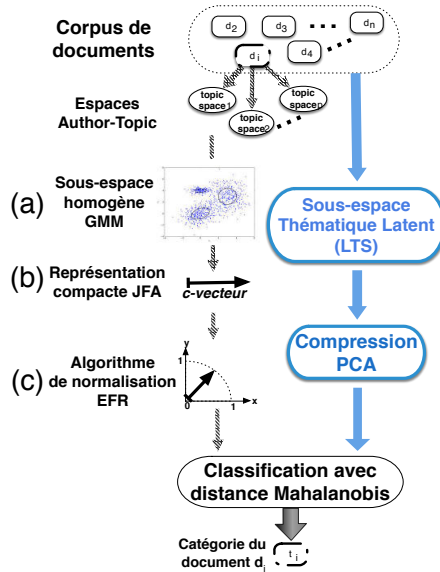


FIGURE 1 – Le sous-espace JFA + GMM ((a)-(c)) et la compression LTS (en bleu).

2.1 Modèle *Author-topic* (AT)

Le modèle AT encode à la fois le contenu du document (la distribution des mots) et les auteurs (distribution des auteurs). Dans le cadre de notre application, un document d est une conversation entre humains agent/utilisateur. L'agent doit annoter cette conversation avec un des 8 thèmes définis en amont. Les thèmes correspondent ici aux "auteurs" dans les modèles AT. Chaque dialogue d est composé d'un ensemble de mots w et un thème a . Dans ce modèle, chaque auteur est associé à une distribution sur les thèmes latents (θ), choisie à l'aide d'une loi de Dirichlet ($\vec{\alpha}$) et à une mixture pondérée afin de sélectionner un thème z . Un mot est alors généré selon la distribution ϕ correspondant au thème z . Cette distribution ϕ est déterminée à partir d'un Dirichlet ($\vec{\beta}$). Par conséquent, ce modèle permet d'encoder les dépendances statistiques entre le contenu du dialogue (les mots w) et le label (thème a) à partir de la distribution du thème latent z dans le dialogue. L'échantillonnage de Gibbs nous permet d'estimer les paramètres du modèle AT afin de représenter un dialogue non observé d avec le r^{eme} espace AT de taille T et d'obtenir un vecteur de paramètres $V_d^{\alpha k} = P(a_k | d)$ relatif à la

représentation du thème d'un dialogue non observé d selon le r^{eme} espace AT Δ_r^n de taille T . Le k^{eme} ($1 \leq k \leq A$) paramètre est :

$$V_{d,r}^{a_k} = \sum_{i=1}^{N_d} \sum_{j=1}^T \theta_{j,a_k}^r \phi_{j,i}^r \quad (1)$$

où A est le nombre de thèmes ; $\theta_{j,a_k}^r = P(a_k|z_j^r)$ est la probabilité du thème a_k à être généré par le thème z_j^r dans le r^{eme} espace de thèmes de taille T . $\phi_{j,i}^r = P(w_i|z_j^r)$ est la probabilité du mot w_i (N_d est la taille du vocabulaire de d) à être généré par le thème z_j^r .

2.2 Représentation c -vecteur

Cette approche, initialement proposée dans (Morchid *et al.*, 2014b), utilise les i -vecteurs pour modéliser la représentation du dialogue à travers chaque espace AT dans un espace homogène. Ces segments courts sont considérés comme des unités de représentation sémantique. Dans notre modèle, le super-vecteur m d'un dialogue d sachant un espace thématique r , est issu de la concaténation des moyennes de chacune des Gaussiennes composant le modèle de GMM (UBM) :

$$\mathbf{m}_{(d,r)} = m + \mathbf{T}\mathbf{x}_{(d,r)} \quad (2)$$

où $\mathbf{x}_{(d,r)}$ contient les coordonnées de la représentation AT du dialogue dans l'espace de variabilité totale réduite ; m est le super-vecteur moyen de l'UBM. \mathbf{T} est la *matrice de variabilité totale* de faible dimension ($MD \times R$), où M est le nombre de Gaussiennes dans l'UBM et D est la taille des caractéristiques. La représentation c -vecteur souffre de 3 problèmes : (i) les c -vecteurs x de l'équation 2 doivent être distribués au sein de la distribution normale $\mathcal{N}(0, I)$, (ii) leur effet "radial" doit être supprimé et (iii) l'espace factoriel de plein rang doit être utilisé pour appliquer des transformations discriminantes. Une solution à ces 3 problèmes (*Eigen Factor Radial* (EFR)) a été développée dans (Bousquet *et al.*, 2011) en standardisant les c -vecteurs comme décrit dans la figure 2.

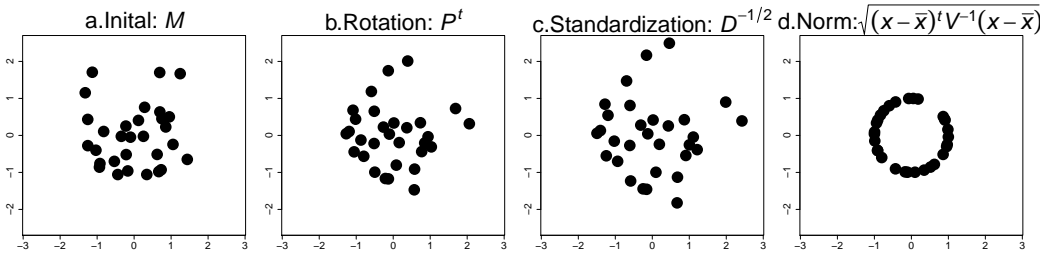


FIGURE 2 – Effet de la standardisation des c -vecteurs à l'aide de l'algorithme EFR.

2.3 Sous-espace Thématique Latent (LTS)

La représentation c -vecteur doit projeter les dialogues dans un UBM-GMM afin d'obtenir un super-vecteur de très grande dimension (la taille de la représentation thématique multipliée par le nombre de Gaussiennes dans l'UBM). Le LTS est composé d'un ensemble d'espaces latents et considère chacun de ces derniers comme un sous-domaine sur lequel est projeté chaque document. Ainsi, toutes les

représentations thématiques d'un document partageant une structure latente commune. Ces paramètres latents partagés définissent le sous-espace thématique latent. Chaque super-vecteur s_d , représentant un document d d'un ensemble de documents de taille N , est partiellement associé à un petit sous-ensemble de caractéristiques latentes et le résidu de cette représentation du document est projeté dans un espace de caractéristiques global partagé par toutes les représentations définissant le sous-espace latent. Le super-vecteur s_d relatif à un dialogue d , est obtenu en concaténant les représentations AT $V_{d,r}^{a_r}$ pour tous les espaces thématiques r . Ainsi, la matrice $\mathbf{S} = [s_0, \dots, s_d, \dots, s_N]$ des super-vecteurs représente les documents dans le LTS. Cette matrice est ensuite compressée au moyen d'une EVD aboutissant à une représentation compacte \mathbf{h}_d dans un espace de faible dimension dont la taille dépend du nombre des valeurs propres e :

$$\mathbf{S} = \mathbf{P}\Delta\mathbf{V}^T \quad (3)$$

où \mathbf{P} est la matrice des vecteurs singuliers à gauche de taille $MD \times N$, \mathbf{V} est la matrice des vecteurs singuliers à droite de taille $N \times N$ ($N \ll MD$) et Δ est la matrice diagonale des valeurs singulières. N est le rang de la matrice \mathbf{S} . Plus d'informations sur l'EVD sont disponibles dans (Abdi & Williams, 2010) et (Golub & Reinsch, 1970). La représentation compacte $\mathbf{h}_{(d,e)}$, de taille e (nombre des valeurs propres), relative à un super-vecteur s_d de \mathbf{S} , est définie comme suit :

$$\mathbf{h}_{(d,e)} = (\mathbf{s}_d - \bar{s}) \cdot \mathbf{V}_e^T \quad (4)$$

où \mathbf{V}_e est la matrice réduite de vecteurs propres relative aux e plus grandes valeurs propres présentes dans la matrice diagonale Δ et \bar{s} est la moyenne de tous les super-vecteurs des documents. En outre, cette représentation compacte d'un document ne nécessite ni l'apprentissage d'un espace commun (comme un UBM-GMM), les espaces thématiques étant un espace de caractéristiques homogène (figure 1-(a) et (b)), ni la normalisation du super-vecteur avec, par exemple, l'algorithme EFR (figure 1-(c)).

3 Protocole Expérimental

L'efficacité de la représentation compacte proposée dans le LTS est évaluée dans le cadre du corpus DECODA (Bechet *et al.*, 2012). Ce corpus comporte environ 74 heures d'enregistrements audio relatifs à 1 514 conversations téléphoniques manuellement annotées en 8 ($A = 8$) thèmes : *Itinéraire*, *Objets trouvés*, *Horaire*, *Carte de transport*, *État du trafic*, *Prix du ticket*, *Infractions*, *Offres spéciales*. 740 conversations servent pour la phase d'apprentissage, 175 pour le développement et 327 pour le test.

La transcription des dialogues a été effectuée par le SRAP Speeral du LIA (Linarès *et al.*, 2007). Les paramètres du modèle acoustique ont été estimés à partir de 150 heures de conversations téléphoniques. Le vocabulaire contient 5 782 mots. Un modèle de langage 3-grammes a été obtenu en adaptant un ML basique sur les transcriptions du corpus d'apprentissage. Ce système atteint un taux d'erreur de mots de 33.8% sur le corpus d'apprentissage, 45.2% sur le corpus de développement, et 49.5% sur celui de test. Ces taux élevés sont dus aux disfluences et aux environnements acoustiques défavorables (par exemple, un appel depuis une rue bruyante avec un téléphone portable).

Une approche de classification fondée sur la distance de Mahalanobis (Morchid *et al.*, 2014a) est utilisée pour trouver le thème principal d'un dialogue. Cette approche probabiliste ignore le processus par lequel les vecteurs ont été extraits. Une fois qu'un vecteur compact a été obtenu à partir d'un

document, le mécanisme de sa représentation est ignoré et il est considéré comme une observation provenant d'un modèle probabiliste génératif. La métrique de Mahalanobis affecte un document d au thème le plus probable C . Sachant un ensemble de documents d'apprentissage, soit \mathbf{W} la matrice de covariance intra-document définie par :

$$\mathbf{W} = \sum_{k=1}^K \frac{n_t}{\bar{N}} \mathbf{W}_k = \frac{1}{n} \sum_{k=1}^K \sum_{i=0}^{n_t} (x_i^k - \bar{x}_k) (x_i^k - \bar{x}_k)^t \quad (5)$$

où \mathbf{W}_k est la matrice de covariance du k^{eme} thème C_k , n_t le nombre de phrases dans le thème C_k , N est le nombre total de documents, et \bar{x}_k la moyenne de tous les documents x_i^k de C_k . Les documents ne contribuent pas de la même manière dans la covariance. C'est pour cette raison que le terme $\frac{n_t}{N}$ est introduit dans l'équation 5. En assumant l'homoscédasticité (égalité entre les classes de covariances) et la Gaussianité de la densité conditionnelle, une nouvelle observation x dans les données de test peut être affectée au thème le plus probable $C_{k_{\text{Bayes}}}$ en utilisant le classifieur s'appuyant sur la loi de décision de Bayes :

$$C_{k_{\text{Bayes}}} = \arg \max_k \left\{ -\frac{1}{2} (x - \bar{x}_k)^t \mathbf{W}^{-1} (x - \bar{x}_k) + a_k \right\} \quad (6)$$

où $a_k = \log(P(C_k))$. Il est à noter que, avec ces hypothèses, l'approche bayésienne est similaire à l'approche géométrique de Fisher. x est affectée à la classe de la plus proche moyenne, selon la métrique de Mahalanobis (Xing *et al.*, 2002) de \mathbf{W}^{-1} :

$$C_{k_{\text{Bayes}}} = \arg \max_k \left\{ -\frac{1}{2} \|x - \bar{x}_k\|_{\mathbf{W}^{-1}}^2 + a_k \right\} \quad (7)$$

4 Expériences et résultats

La section 4.1 présente les résultats obtenus avec deux représentations fondées sur les modèles AT et la technique c -vecteur. Ensuite, la proposition compacte originale s'appuyant sur un LTS est comparée avec la représentation c -vecteur dans la section 4.2.

4.1 Impact de la compression c -vecteur

Les expériences sont conduites en utilisant 500 espaces AT en faisant varier le nombre de thèmes de 2 à 505 par pas de 1. Une solution classique consiste à chercher l'espace thématique qui atteint la meilleure performance. La figure 3 présente les performances de classification de thèmes obtenues sur le corpus de développement (figure 3) et sur le corpus de test (figure 3-(b)) en utilisant diverses configurations de la représentation AT (*baseline*).

Premièrement, nous pouvons constater que la méthode *baseline* atteint une précision de 86,3 % et 83,8 % respectivement sur les données de développement et de test. Cependant, nous remarquons que la performance de classification est plutôt instable et peut varier complètement d'un espace thématique à l'autre. L'écart entre les bornes inférieures et supérieures des résultats de classification est aussi important (21,2 points). Par conséquent, il est crucial de trouver la meilleure configuration d'espace AT pour cette tâche de classification. En utilisant les paramètres du point de fonctionnement

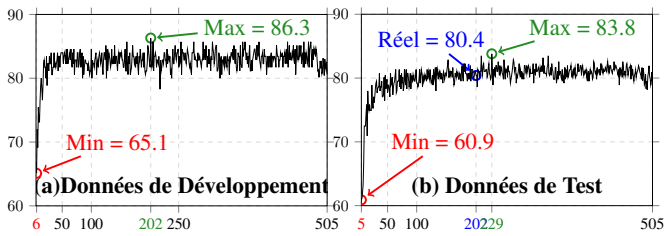


FIGURE 3 – La précision de la classification en thèmes (%) en utilisant diverses représentations thématiques sur les corpus de développement et de test avec différentes configurations expérimentales. L’axe X représente le nombre n de classes composant l’espace thématique ($5 \leq n \leq 505$).

($n = 202$ thèmes) estimé sur le corpus de développement, la classification atteint une précision de 80,4 %.

Le tableau 1 présente la méthode c -vecteur associée à l’algorithme de normalisation EFR (Morchid *et al.*, 2014a). Nous pouvons premièrement remarquer que cette représentation compacte permet de surpasser la performance de la configuration d’espace AT optimale avec un gain de 1,9 points sur les données de test. L’incohérence de la performance de classification n’est pas observée avec cette approche (comme déjà mentionné dans (Morchid *et al.*, 2014a)). En effet, la configuration qui a obtenu la meilleure précision sur les données de développement est aussi optimale pour les données de test. En outre, l’écart de performance entre les différentes configurations du c -vecteur est beaucoup moins faible. la précision minimale atteinte sur les données de test est de 78,9 % contre 60,9 % avec les espaces AT (voir figure 3-(b)).

taille du c -vecteur	DEV			TEST		
	Nb gaussiennes dans le GMM-UBM					
	32	64	128	32	64	128
80	80,6	82,3	83,1	79,2	81,0	80,4
100	81,7	84,6	83,1	78,9	82,3	80,4
120	84,0	81,7	82,3	80,4	79,2	81,8

TABLE 1 – Précision de classification en thèmes (%) dans l’espace de variabilité totale avec différentes tailles de l’UBM et des c -vecteurs

4.2 Impact de la compression LTS

Les résultats obtenus en utilisant la représentation LTS sont présentés dans la figure 4. Afin de comparer les performances des différentes approches (AT/ c -vecteurs/LTS), les meilleurs résultats sont présentés dans le tableau 2. Il faut préciser que ces résultats sont obtenus dans des conditions d’application “réelles”, la configuration optimale (nombre de thèmes dans l’espace thématique) étant choisie dans la phase de développement. En conséquence, il pourrait exister un meilleur point de fonctionnement pour le corpus de test, ce qui explique l’écart entre les résultats répertoriés dans le tableau 2 et ceux des figures 3 et 4. Nous remarquons que la représentation LTS surpasse chacune des représentations AT (baseline) et c -vecteur sur le corpus de développement (figure 4-(a)) et de test

(figure 4-(b)) avec respectivement une précision de 89,7 % (+2, 7 points) et 85,3 % (+4, 6 points). En outre, les courbes du modèle LTS sont également plus stables et plus robustes que celles de la représentation c -vecteur. En effet, l'écart entre les deux valeurs extrêmes, qui est de 3,3 points avec les c -vecteurs, est égal à 2,1 points pour la représentation LTS sur les données de test.

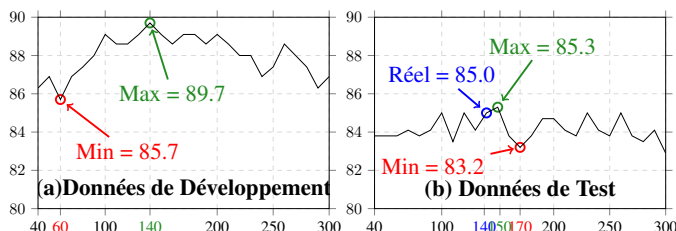


FIGURE 4 – Précisions de la classification en thèmes (%) en utilisant un vecteur compact de représentations LTS sur les corpus de développement et de test avec différentes configurations expérimentales. L'axe X représente le nombre des valeurs propres m ($40 \leq m \leq 300$).

Approche de représentation	DEV		TEST
	taille	prec. %	prec. %
Modèle AT (baseline)	202	86,3	80,4
Modèle AT + c -vecteur	100	84,6	82,3
Modèle ATAT + LTS	140	89,7	85,0

TABLE 2 – Précision de la classification en thèmes (%) sur le corpus de test en utilisant la configuration optimale obtenue sur le corpus de développement.

5 Conclusion

La performance des SRAP dépend des conditions d'enregistrement et la qualité des transcriptions automatiques produites influe sur les tâches de compréhension du langage. Ce papier propose une solution pour faire face aux erreurs de transcription en projetant un dialogue dans un sous-espace de caractéristiques robustes appelé Sous-espace Thématique Latent (LTS). Les expériences conduites dans le cadre de la classification de conversations a montré l'efficacité du modèle LTS proposé par rapport à l'utilisation des représentations c -vecteur et AT classiques. Cette représentation haut-niveau permet d'améliorer considérablement la performance de la tâche d'identification de thèmes avec un gain supérieur à 3 et 2 points respectivement par rapport à l'utilisation des représentations AT et c -vecteur. Le modèle LTS est combiné avec une compression PCA vu la faible taille des données (740 dialogues dans le corpus d'apprentissage). Dans la continuité de ce travail, il est intéressant d'évaluer cette représentation prometteuse avec des données plus volumineuses que celles du projet DECODA. Ainsi, d'autres méthodes de compression comme l'auto-encodeur et la PCA Probabiliste pourraient obtenir des résultats meilleurs durant les différentes tâches d'analyse de la parole.

Références

- ABDI H. & WILLIAMS L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews : Computational Statistics*, **2**(4), 433–459.
- BECHET F., MAZA B., BIGOUROUX N., BAZILLON T., EL-BEZE M., DE MORI R. & ARBILLOT E. (2012). : LREC'12.
- BLEI D. M., NG A. Y. & JORDAN M. I. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research*, **3**, 993–1022.
- BOUSQUET P.-M., MATROUF D. & BONASTRE J.-F. (2011). Intersession compensation and scoring methods in the i-vectors space for speaker recognition. In *Interspeech*, p. 485–488.
- DEHAK N., KENNY P. J., DEHAK R., DUMOUCHEL P. & OUELLET P. (2011). Front-end factor analysis for speaker verification. *IEEE Transactions on Audio, Speech, and Language Processing*, **19**(4), 788–798.
- EISENSTEIN J. & BARZILAY R. (2008). Bayesian unsupervised topic segmentation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, p. 334–343 : ACL.
- GOLUB G. H. & REINSCH C. (1970). Singular value decomposition and least squares solutions. *Numerische mathematik*, **14**(5), 403–420.
- HAZEN T. (2011). Topic identification. *Spoken Language Understanding : Systems for Extracting Semantic Information from Speech*, p. 319–356.
- LAGUS K. & KUUSISTO J. (2002). Topic identification in natural language dialogues using neural networks. In *Proceedings of the Third SIGdial Workshop on Discourse and Dialogue*, p. 95–102, Philadelphia, Pennsylvania, USA : Association for Computational Linguistics.
- LINARÈS G., NOCÉRA P., MASSONIE D. & MATROUF D. (2007). The lia speech recognition system : from 10xrt to 1xrt. In *Text, Speech and Dialogue*, p. 302–308 : Springer.
- MAY C., FERRARO F., MCCREE A., WINTRODE J., GARCIA-ROMERO D. & VAN DURME B. (2015). Topic identification and discovery on text and speech.
- MELAMED I. & GILBERT M. (2011). Speech analytics. *Spoken Language Understanding : Systems for Extracting Semantic Information from Speech*, p. 397–416.
- MORCHID M., BOUALLEGUE M., DUFOUR R., LINARÈS G., MATROUF D. & DE MORI R. (2014a). I-vector based approach to compact multi-granularity topic spaces representation of textual documents. In *EMNLP : SIGDAT*.
- MORCHID M., BOUALLEGUE M., DUFOUR R., LINARÈS G., MATROUF D. & DE MORI R. (2014b). I-vector based representation of highly imperfect automatic transcriptions. In *Conference of the International Speech Communication Association (Interspeech) 2014 : ISCA*.
- MORCHID M., DUFOUR R., BOUALLEGUE M. & LINARÈS G. (2014c). Author-topic based representation of call-center conversations. In *SLT : IEEE*.
- PURVER M. (2011). Topic segmentation. *Spoken Language Understanding : Systems for Extracting Semantic Information from Speech*, p. 291–317.
- ROSEN-ZVI M., GRIFFITHS T., STEYVERS M. & SMYTH P. (2004). The author-topic model for authors and documents. In *Uncertainty in artificial intelligence*, p. 487–494 : AUAI Press.
- TUR G. & DE MORI R. (2011). *Spoken language understanding : Systems for extracting semantic information from speech*. John Wiley & Sons.
- XING E. P., JORDAN M. I., RUSSELL S. & NG A. (2002). Distance metric learning with application to clustering with side-information. In *Advances in neural information processing systems*.