Segmentation et classification non supervisée de conversations téléphoniques automatiquement retranscrites

Laurent Bozzi, Philippe Suignard, Claire Waast-Richard

EDF R&D
1, avenue du Général de Gaulle
92141 Clamart Cedex
Laurent.Bozzi@edf.fr, Philippe.Suignard@edf.fr, Claire.Waast-Richard@edf.fr

Résumé : Cette étude porte sur l'analyse de conversations entre des clients et des télé-conseillers d'EDF. Elle propose une chaîne de traitements permettant d'automatiser la détection des sujets abordés dans chaque conversation. L'aspect multi-thématique des conversations nous incite à trouver une unité de documents entre le simple tour de parole et la conversation entière. Cette démarche enchaîne une étape de segmentation de la conversation en thèmes homogènes basée sur la notion de cohésion lexicale, puis une étape de text-mining comportant une analyse linguistique enrichie d'un vocabulaire métier spécifique à EDF, et enfin une classification non supervisée des segments obtenus. Plusieurs algorithmes de segmentation ont été évalués sur un corpus de test, segmenté et annoté manuellement : le plus « proche » de la segmentation de référence est C99. Cette démarche, appliquée à la fois sur un corpus de conversations transcrites à la main, et sur les mêmes conversations décodées par un moteur de reconnaissance vocale, aboutit quasiment à l'obtention des 20 mêmes classes thématiques.

Abstract : This study focuses on the analysis of conversations and between clients and EDF agent. It offers a range of treatments designed to automate the detection of the topics covered in each conversation. As the conversations are multi-thematic we have to find a document unit, between the simple turn of speech and the whole conversation. The proposed approach starts with a step of segmentation of the conversation (based on lexical cohesion), and then a stage of text-mining, including a language enriched by a vocabulary specific to EDF, and finally a clustering of the segments. Several segmentation algorithms were tested on a test corpus, manually annotated and segmented : the "closest" to the reference segmentation is C99. This approach, applied to both a corpus of conversations transcribed manually, and on the same conversations decoded by a voice recognition engine, leads to almost obtain the same 200 clusters.

Mots-clés : audio-mining, text mining, segmentation, classification, catégorisation, reconnaissance vocale, données textuelles, conversations téléphoniques, centre d'appel.

Keywords : audio-mining, text mining, segmentation, clustering, categorization, voice recognition, textual data, phone conversation, call center.

1 Introduction

Le développement des NTIC dans une entreprise de la taille du Groupe EDF a entraîné une augmentation du volume et des flux de données hétérogènes (téléphone, internet …) qui peuvent être traitées par des méthodes de fouille et de classement automatique. En particulier en ce qui concerne les centres d'appels, les directions marketing cherchent à mieux connaître les motifs d'appels des clients, à en faire une typologie. Il s'agit de comprendre la relation des clients à l'entreprise mais aussi de mesurer les impacts de certains événements (publicité, nouvelles offres…). Le projet Infom@gic, du pôle de compétitivité CAP DIGITAL, et le sous-projet CallSurf s'inscrivent dans le cadre du traitement automatique de conversations.

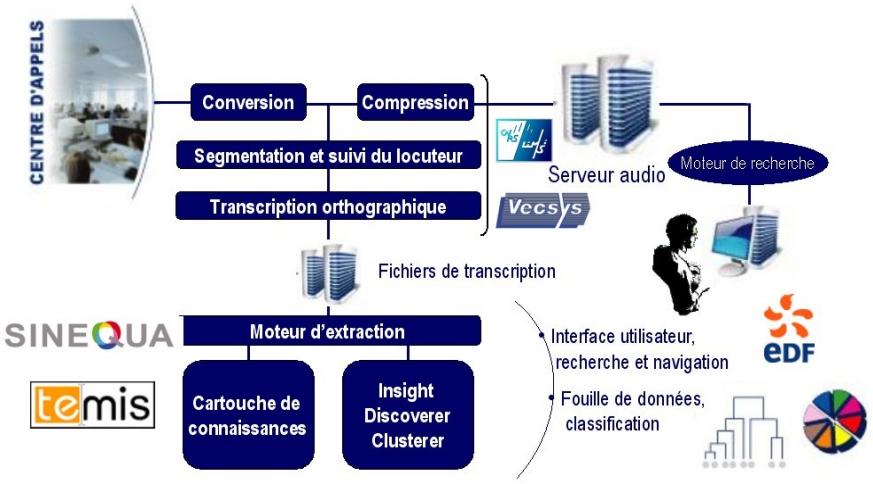

Figure 1 : Schéma fonctionnel de CallSurf

CallSurf (Garnier-Rizet, 2008) est un système de transcription et de classement de données audio comportant plusieurs étapes : enregistrement des conversations entre client et télé-conseiller dans un centre d'appel, transcription automatique des conversations en texte, classification des conversations préalablement segmentées en thèmes homogènes en un ou plusieurs thèmes et catégorisation qui consiste à calculer un modèle d'affectation d'une nouvelle conversation à un ou plusieurs des thèmes prédéfinis. Les sections suivantes décrivent les premiers travaux ayant conduit à une classification automatique satisfaisante des conversations téléphoniques.

2 Segmentation des conversations

Une classification des conversations dans leur ensemble s'avère trop grossière car une conversation peut comporter plusieurs thèmes. A l'inverse, classifier les tours de paroles est difficile car beaucoup d'échanges oraux sont peu porteurs de sens. C'est pourquoi, on cherche d'abord à découper les conversations en segments homogènes pour ensuite les classifier. La suite de cette partie présente la démarche suivie afin d'aboutir à la meilleure segmentation possible : constitution d'un corpus de référence, test et choix d'algorithmes.

2.1 Corpus de référence

Plusieurs conversations ont été segmentées à la main par des « experts » EDF aboutissant à une nomenclature de segmentation et d'annotation en thèmes (avec environ une vingtaine de thèmes possibles comme ouverture, clôture, motif, facture…). Cette description a ainsi permis

de segmenter et d'annoter manuellement environ 150 conversations, représentant environ 15h, et ainsi constituer un corpus de référence. Ce corpus comporte des particularités par rapport à d'autres corpus utilisés pour tester des algorithmes de segmentation (DEFT par exemple) : il s'agit d'un corpus de parole conversationnelle, donc assez différent du texte écrit, avec des dysfluences, reprises, répétitions, mots tronqués… entre deux ou plusieurs personnes, avec du vouvoiement entre client et agent EDF, mais tutoiement entre agents. Par ailleurs, il présente une grande variabilité sur la longueur des segments.

2.2 Algorithmes testés

Les algorithmes testés ici font partie de la famille des méthodes non supervisées. Elles s'appuient sur la notion de cohésion lexicale et partent du principe qu'à un thème donné, correspond un vocabulaire spécifique. Ainsi, pour détecter les changements de thème, il suffit de détecter les ruptures lexicales. La méthode fondatrice du domaine s'appelle « TextTiling » (Hearst, 1997). Pour nos conversations composées de tours de parole, elle consiste à considérer deux fenêtres lexicales constituées de N tours de paroles situés avant et après le tour de parole i. Pour chaque tour de parole i, on compare le vocabulaire des deux fenêtres par la mesure des cosinus. En faisant glisser les deux fenêtres le long des tours de parole, on obtient une courbe de cohésion. Détecter les chutes de cette cohésion permet de détecter les ruptures de thème.

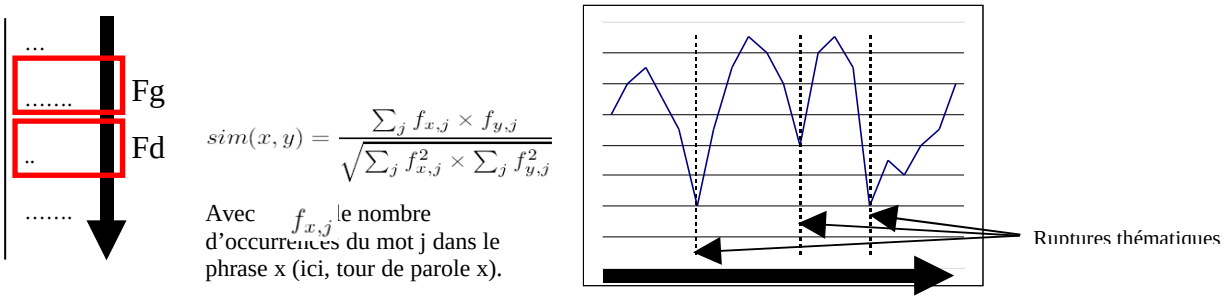

Figure 2 : Principe de l'algorithme TextTiling

« C99 » (Choï, 2000), est la grande référence à laquelle toutes les autres méthodes se comparent. Cette méthode commence par calculer les similarités 2 à 2 entre tous les unités textuelles (ici les tours de parole), puis regroupe ces unités, de manière successive, jusqu'à obtenir la répartition obtenant la meilleure densité. Cet algorithme a été utilisé tel quel, directement téléchargé depuis le site de son auteur.

Une méthode fortement inspirée de « SegGen » (Lamprier, 2007) a été reprogrammée : elle consiste à générer aléatoirement une population de N segmentations pour un texte initial donné. Par croisements successifs des segmentations entre elles, elle converge vers la « meilleure » segmentation en maximisant un score « intra segment » qui mesure la cohésion du segment lui-même et en minimisant un score « inter segment » mesurant la similarité entre deux segments consécutifs.

2.3 Mesures de distance

Pour tester la qualité des segmentations obtenues par ces algorithmes, nous avons retenu la mesure WindowDiff (WD) (Pevzner & Hearst, 2002) :

$$WindowDiff(ref, hyp) = \frac{1}{N-k} \sum (|b(ref_i, ref_{i+k}) - b(hyp_i, hyp_{i+k})|)$$

Avec *b*(xi, xj) le nombre de frontières entre les tours i et j dans le texte x, contenant N tours de parole, k étant un entier fixé au départ (par exemple la moitié de la longueur moyenne des segments du document de référence).

2.4 Tests et résultats

Chacun des 3 algorithmes a été testé sur les 150 conversations en faisant varier les traitements en amont sur le texte : suppression de la ponctuation, suppression de mots a priori parasites comme euh, hum, hein, ben… ou inutiles comme oui, non, bon, d'accord… ou encore en racinisant les mots (pour supprimer les suffixes flexionnels et dérivationnels). Les conclusions de tous ces tests sont les suivantes :

- C99 est retenu pour générer la segmentation : avec un WD moyen de 0.45 et un nombre moyen de 7.3 tours de parole par segment contre 8.3 dans le corpus de référence. Il présente l'intérêt d'être « tout terrain » (même quand il ne fournit pas toujours les meilleurs résultats, la segmentation produite est « correcte »).
- Tous ces tests poussent à l'interrogation concernant la fiabilité de la mesure WD[1]. Par exemple, deux segmentations différentes peuvent obtenir un même score, mais avec un nombre de segments très différents (10 segments dans un cas et 30 dans l'autre).
- Il faut également s'interroger sur les segmentations de référence (il peut y avoir jusqu'à 31 segments dans une conversation). Pour cela, d'autres conversations sont en train d'être segmentées et annotées à la main pour augmenter la taille du corpus de référence et par des personnes différentes.

3 Classification des conversations

La partie de classification des conversations s'est faite en plusieurs étapes, et sur plusieurs jeux de données, 170 heures (1421 conversations) retranscrites manuellement, dont 20 heures (188 conversations) de façon détaillée.

3.1 Analyse sur les données retranscrites manuellement

Pour les premiers essais, la démarche d'analyse textuelle suivante a été suivie : analyse morpho-syntaxique, lemmatisation, transformation des données textuelles en matrice termes*documents, réduction de cette matrice par décomposition en SVD[2] , pondération par la méthode tf*idf, classification des vecteurs représentant x% de l'inertie totale (x étant paramétrable) par l'algorithme EM (Expectation-Maximization).

En classifiant les tours de parole en 10 classes, on s'est aperçus de la nécessité de segmenter les conversations (cf. §2). Pour les essais suivants, et afin d'optimiser les résultats, nous avons joué sur plusieurs leviers :

- Segmentation de la conversation en parties homogènes (cf. partie précédente).
- Amélioration de la partie linguistique, en utilisant conjointement les possibilités offertes par Temis pour développer une « cartouche » spécifique à EDF[3], et l'analyseur morpho-syntaxique de Sinequa, adapté lui aussi aux données d'EDF.

[1] Une autre mesure, δ-Front (Fernandez et al., 2008) a également été testée, mais présente le même genre d'inconvénients.

[2] SVD : Singular Value Decomposition (Décomposition en valeurs singulières)

- Classification, par la méthode k-means. Au vu du nombre d'essais à pratiquer, sa rapidité d'exécution était un plus, pour des résultats très proches, au niveau qualitatif, de ceux obtenus avec l'algorithme EM. Afin de réduire la matrice termes*documents, nous avons essentiellement conservé les mots présents au moins 3 fois dans l'ensemble du corpus, et étiquetés comme verbes, noms communs, adjectifs et concepts créés à l'étape précédente.
- Augmentation du nombre de classes : on a choisi de faire une classification « à plat » avec beaucoup de classes, plutôt qu'une classification hiérarchique, avec moins de classes au départ, elles-mêmes sous-classifiées.

Cette méthode nous a permis de mettre en évidence une vingtaine de classes, certaines portant sur la structure de la conversation, et d'autres sur son thème : clôture/ouverture de conversation, paiement de facture / mensualisation, coordonnées / références client, règlement facture / problème de paiement, résiliation, coordonnées de l'entreprise, coordonnées du client, prise de rendez-vous, problème de puissance (souvent demande d'augmentation), mise en service, demande de branchement / raccordement, coordonnées bancaires, usages chauffage, climatisation et autres, énergies concurrentes (gaz), mise en attente de la part de l'agent / recherche du dossier client, dates ou jours de rendez-vous. En revanche, la taille des classes n'est pas homogène, un quart des segments de conversations concernent soit l'ouverture soit la clôture de la conversation (car ces classes structurelles font partie de quasiment toutes les conversations). De même, les échanges de coordonnées (référence client, adresse, téléphone, etc.) représentent une bonne part des segments, 20%.

3.2 Analyse sur les données retranscrites automatiquement

Notre objectif étant de proposer une chaîne de traitements complètement automatisée, on a alors utilisé un premier jeu de données retranscrites automatiquement (reconnaissance vocale). Les résultats des tests suivants sont toutefois provisoires, car ces données ont servi de base d'apprentissage au modèle de langage.

On a constaté que l'outil de classification était assez robuste, puisqu'on retrouvait les principaux thèmes des classifications précédentes, sur données transcrites manuellement. Pour illustrer cela, si l'on prend une classe qui nous intéresse particulièrement, comme les « moyens de paiement », on constate que cette classe représente 7% des segments sur les conversations retranscrites manuellement et 6% sur les conversations automatiques. En outre, le champ lexical caractérisant les classes des deux jeux de données est quasiment le même, les mots-clés caractérisant cette classes étant les suivants :

- Sur les données manuelles : *moyen_paiement ; envoyer ; facture ; règlement ; recevoir ; carte_bancaire ; régler ; courrier ; prelevement ; payer ; noter ; relance ; probleme ; mettre ; voir*
- Sur les données automatiques : *moyen_paiement ; envoyer ; facture ; être ; recevoir ; règlement ; courrier ; payer ; régler ; y_avoir ; d'accord ; voir ; prelevement ; probleme ; mettre ; relance ; retard ; attendre ; jour ; septembre*

4 Conclusions et perspectives

Nous proposons donc une chaîne de traitements assez complexe, permettant de passer d'un signal audio, à une liste de thèmes contenus dans ce signal. La détection des thèmes de la

[3] Une « cartouche » de Temis, permet de prendre en compte des expressions régulières, de synonymiser certains mots ou groupes de mots, et surtout de créer des concepts clés spécifiques au vocabulaire employé dans les conversations de télé-conseillers. Par exemple, regroupement sous le concept « moyen de paiement » des termes « chèques », « TIP », « RIB ».

conversation a été grandement facilité par la segmentation thématique automatique (après le choix de C99) et la prise en compte de règles linguistiques propre au vocabulaire d'EDF. Par ailleurs, le passage de données manuelles à des données automatiquement retranscrites n'a pas perturbé ce processus. Cependant, de nouvelles analyses seront réalisées sur des données n'ayant pas servi à caler le modèle de langage.

Plusieurs limites ou problèmes sont apparus. Concernant la segmentation, les algorithmes automatiques présentent plusieurs paramètres. De plus, le choix de l'algorithme de segmentation n'est pas tranché, en raison de la mesure WD utilisée, peut-être pas adaptée à un corpus oral. Enfin, lors de la classification, le nombre de classes doit se faire de manière experte à la suite de plusieurs itérations. De ce fait, à chaque étape (segmentation, classification), nous sommes confrontés à de nombreux choix successifs, dont les effets sur le résultat final sont difficiles à évaluer. Afin d'améliorer les résultats, nous essaierons de régler de façon plus fine les paramètres des algorithmes de segmentation, et tenterons d'améliorer les classifications par des techniques de ré-échantillonnage.

Remerciements

CapDigital, les partenaires du projet :Vecsys, LIMSI, Sinéqua et Témis.

Références

Cailliau F., Poudat C., (2008) Caractérisation lexicale des contributions clients agents dans un corpus de conversations téléphoniques retranscrites, JADT 2008.

Choi F. Y. Y. (2000). Advances in domain independent linear text segmentation. In Proceedings of the 1st Meeting of the North American Chapter of the Association for Computational Linguistics, USA.

Fernandez S., Sanjuan E., Torres-Moreno J-M. (2008). Enertex: un système basé sur l'énergie textuelle. Actes de *Traitement Automatique de la Langue Naturelle,* TALN 2008.

Garnier-Rizet M., Adda G., Cailliau F., Gauvain J. L., Guillemin-Lanne S., Lamel L., Vanni S., Waast-Richard C. (2008). CallSurf - Automatic transcription, indexing and structuration of call center conversational speech for knowledge extraction and query by content, LREC 2008.

Hearst M. A. (1997). Text-tiling : segmenting text into multi-paragraph subtopic passages. *Computational Linguistics*, p. 59–66.

Labadié A., Prince V. (2008). Comparaison de méthodes lexicales et syntaxico-sémantiques dans la segmentation thématique de texte non supervisée. Actes de *Traitement Automatique de la Langue Naturelle,* TALN 2008.

Lamprier S., Amghar T., Levrat B., Saubion F. (2007) SegGen: A Genetic Algorithm for Linear Text Segmentation. Actes de IJCAI.

Pevzner L. & Hearst M. A. (2002). A critique and improvement of an evaluation metric for text segmentation. *Computational Linguistics*, p. 19–36.