

Analyse sémantique de transcriptions automatiques d'appels téléphoniques en français

Emmanuelle Dusserre Ruslan Kalitvianski Mathieu Ruhlmann Muntsa Padró
Eloquent, 5 allée de Palestine, 38610 Gières, France
{prenom.nom}@eloquant.com

RÉSUMÉ

Dans cet article, nous présentons la mise en œuvre d'une chaîne de traitement sémantique complète dédiée aux conversations audio issues de centres d'appel téléphoniques, depuis la phase de transcription automatique jusqu'à l'exploitation des résultats, en passant par l'étape d'analyse sémantique des énoncés. Nous décrivons ici le fonctionnement des différentes analyses que notre équipe développe, ainsi que la plateforme interactive permettant de restituer les résultats agrégés de toutes les conversations analysées.

ABSTRACT

Semantic analysis of automatic phone call transcriptions in French

In this article, we present the implementation of a complete semantic processing chain dedicated to call center phone conversations, from the speech-to-text phase to the exploitation of the results, including the semantic analysis of the utterances. Here we describe the workings of the various analyses that our team develops, as well as the interactive platform that displays the aggregated results of all the analyzed conversations.

MOTS-CLÉS : analyse sémantique, centres d'appels, transcriptions automatiques, *speech-to-text*
KEYWORDS: semantic analysis, speech-to-text, call centers

1 Introduction

Notre société propose une gamme complète de logiciels pour la gestion de la relation client. Nous proposons, entre autres, des logiciels pour la gestion des centres d'appel, pour la réalisation d'enquêtes de satisfaction, ainsi que des modules pour faire l'analyse sémantique du texte. L'un des projets les plus innovants de notre entreprise est le développement des modules d'analyse sémantique des transcriptions automatiques de conversations téléphoniques, modules qui sont clés pour une meilleure exploitation des données audios collectées par nos clients.

Cet article décrit, d'une part, la plateforme en ligne qui restitue de manière graphique et interactive les résultats des analyses sémantiques quotidiennes et, d'autre part, le fonctionnement du moteur d'analyse sémantique sous-jacent et de ses modules. L'analyse s'appuie sur des transcriptions automatiques réalisées par un partenaire, qui effectue également une diarisation.

La restitution présentée dans cette démonstration est destinée aux chefs de projet des services de consommateurs de notre client, qui désirent avoir un panorama synthétique et interactif des interactions téléphoniques entre les agents de support de leur entreprise et leurs clients.

2 L'analyse sémantique : briques et mode opératoire

Nous traitons de manière quotidienne les audios transmises par notre client de façon sécurisée, datant de la veille. Les audios sont au format `wav`, en monocanal, 8 KHz, 64 Kbit/seconde, encodées en G.711. Ces données sont accompagnées d'un fichier de métadonnées qui renseigne des informations utiles au client au moment de la restitution des analyses (date et heure, pôle, etc.).

Les audios sont envoyées à la transcription, réalisée par un partenaire, qui retourne des documents XML contenant, d'une part une information sur le nombre et le genre des locuteurs, et d'autre part, la transcription segmentée en mots chronodatés, regroupés en énoncés étiquetés par l'identifiant de leur locuteur. Ici les énoncés sont des segments de parole délimités soit par des silences, soit par les énoncés d'un autre locuteur.

Ces transcriptions et les métadonnées des audios sont converties vers la structure de données unique sur laquelle opèrent l'ensemble de nos modules d'analyse sémantique, qui sont ordonnés en chaîne de traitement *pipeline*. Chaque conversation est analysée de façon indépendante des autres conversations, et l'unité d'analyse est l'énoncé dans la plupart des cas.

Nos modules d'analyse sémantique s'appuient sur une analyse morphosyntaxique réalisée avec Talismane (Urieli & Tanguy, 2013), et incluent les analyses sémantiques suivantes, réalisées dans cet ordre :

- Une classification des locuteurs (identifiés en amont par la diarisation), en « agent » ou « client ». Cela facilite la lecture de la transcription et permet de filtrer les énoncés dans l'interface de restitution décrite dans la section 3. Cette classification est effectuée sur les ensembles des énoncés produits par chaque locuteur. Elle est réalisée par un classifieur binaire MaxEnt (Nigam et al., 1999) entraîné avec des n-grammes extraits de 200 conversations annotées manuellement et choisies sur la base de leurs durées (afin de capter les différences lexicales observées entre les conversations longues, moyennes et courtes).
- Une anonymisation des noms, lieux, numéros (tout en évitant d'anonymiser les numéros de références produit), adresses physiques et électroniques. Cela garantit un niveau de confidentialité qui nous permet contractuellement de conserver les transcriptions analysées pour une durée d'un an. Cette anonymisation est réalisée à l'aide de listes de prénoms, noms, noms de villes et régions, et de plusieurs dizaines d'expressions régulières étalonnées sur environ 1200 énoncés contenant diverses informations personnelles.
- Une extraction de concepts récurrents : il s'agit de noms ou groupes nominaux significatifs extraits selon le modèle proposé par (Sclano & Velardi, 2007) qui compare les fréquences d'apparition des termes entre le corpus spécifique et un corpus généraliste.
- Une extraction d'entités nommées du domaine de notre client, fondée sur une liste de termes établis avec le client et de leurs synonymes. Nous avons constitué différents gazetteers, spécifiques aux entités nommées du secteur de notre client, afin de relever toutes les occurrences de ces termes dans les conversations. Les gazetteers suivent une syntaxe relativement simple : le terme, qui peut s'écrire sous forme d'expression régulière, et le tag sémantique qu'on souhaite lui attribuer. Par exemple, le nom d'un produit a pour tag sémantique la famille du produit à laquelle il se réfère.
- Un étiquetage multi-label de chaque énoncé par un sous-ensemble d'une vingtaine de catégories pertinentes pour le métier de notre client, elles-mêmes groupées en une dizaine de méta-catégories. Ce traitement est fondé sur l'apprentissage automatique (utilisant comme traits les mots, leurs lemmes, les groupes nominaux, et les annotations sémantiques réalisées en amont via notamment des listes de gazetteers propres au domaine), et sur des règles expertes exprimées

à l'aide du formalisme TokensRegex (Chang & Manning, 2014). Ces dernières nous permettent d'une part d'enrichir sémantiquement les données destinées à la classification en annotant les expressions polylexicales et, d'autre part, de corriger les catégories là où l'apprentissage machine rencontre des difficultés. Pour ce second cas, les TokensRegex ajustent les scores attribués par le classifieur lorsqu'elles rencontrent une expression polylexicale, en revalorisant le score sur la catégorie à laquelle l'expression appartient.

- Un étiquetage multi-label de la conversation complète, qui agrège d'une façon particulière l'ensemble des étiquettes attribuées aux énoncés de la conversation, en suivant les règles de dominance thématique spécifiées par notre client.
- Une extraction des fragments exprimant une opinion positive ou négative. Elle s'appuie sur des règles expertes TokensRegex développées spécifiquement pour les audios, car nous nous sommes aperçus que nos grammaires d'opinions développées pour les verbatims écrits étaient inadaptées pour les transcriptions de conversations : outre les erreurs de transcriptions, le langage parlé diffère fortement du langage écrit, que ce soit par sa structure rhétorique et discursive ou par le registre et les choix lexicaux.

La *pipeline* produit en sortie, pour chaque conversation, une sérialisation XML de l'ensemble des énoncés enrichis de leurs métadonnées et de leurs analyses. Ces fichiers XML sont ensuite indexés sur un serveur Elasticsearch¹ et restitués dans une instance de la plateforme interactive Kibana², décrite ci-dessous.

3 L'interface exploratoire des résultats

L'interface web de la plateforme restitue d'une part les métadonnées des audios (volumétrie, durée moyenne, source) et, d'autre part, les résultats de l'analyse sémantique via un tableau de bord qui inclut une quinzaine de visualisations interactives sous forme d'histogrammes, de « camemberts » ou de nuages de mots (cf. Figure 1), groupées en cinq sections : « Métadonnées », « Concepts et Produits », « Catégorisation », « Opinions » et « Conversations ». En cliquant sur les éléments d'une visualisation un filtre s'applique pour mettre à jour l'ensemble du tableau de bord.

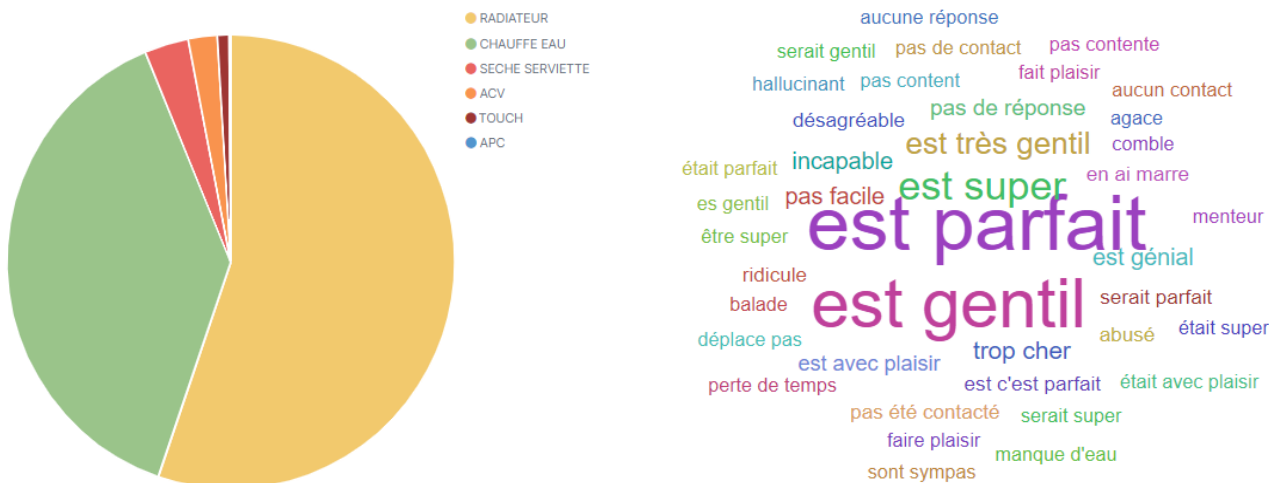


FIGURE 1 : illustration d'un « camembert » des pôles produits abordés dans les conversations, et d'une détection de fragments portant une polarité positive ou négative.

¹ <https://www.elastic.co/fr/elasticsearch>

² <https://www.elastic.co/fr/kibana>

Les visualisations servent à afficher de façon graphique la fréquence et/ou la proportion d'éléments relevés. Ainsi la taille des mots ou des expressions dans les nuages de mots correspondent à leurs fréquences relatives.

La Figure 2 ci-dessous illustre un prototype d'affichage des conversations transcrites et analysées.



FIGURE 2 : illustration du prototype du lecteur audio synchronisé avec la transcription analysée. Les pastilles colorées et la coloration des énoncés représentent les catégories attribuées aux énoncés et leurs occurrences dans la conversation. Actuellement cet affichage n'illustre pas la détection des opinions, ni l'aspect multi-label de la catégorisation.

Remerciements

Nous voudrions remercier David Graceffa, Tim Robert, Mathieu Bliem, Daniel Chipan et Hui Pillot pour leur travail sur le prototype du lecteur audio présenté ci-dessus, ainsi que Myriam El Helou pour sa participation aux modules d'analyse sémantique. Nous remercions également le comité de relecture pour leurs critiques constructives.

Références

- CHANG A. X., & MANNING C. D. (2014). TokensRegex: Defining cascaded regular expressions over tokens. *Stanford University Computer Science Technical Reports. CSTR 2*: 2014.
- NIGAM K., LAFFERTY J., & MCCALLUM A. (1999). Using maximum entropy for text classification. *In IJCAI-99 workshop on machine learning for information filtering (p. 61-67)*. 1999.
- SCLANO, F., ET VELARDI, P. (2007). TermExtractor: a Web Application to Learn the Shared Terminology of Emergent Web Communities. *In R. J. Gonçalves, J. P. Müller, K. Mertins, et M. Zelm (Éd.), Enterprise Interoperability II (p. 287-290)*, 2007. Springer London.
- URIELI A. & TANGUY L. (2013). L'apport du faisceau dans l'analyse syntaxique en dépendances par transitions : études de cas avec l'analyseur Talismane. *20e conférence du Traitement Automatique du Langage Naturel (TALN)*, 2013, Sables d'Olonne, France.