

FMO : un outil d'analyse automatique de l'opinion

Jean-Leon Bouraoui Marc Canitrot

Prometil, 42 Avenue du Général de Croutte, 31100 Toulouse

{jl.bouraoui,m.canitrot}@prometil.fr

RÉSUMÉ

Nous décrivons notre prototype d'analyse automatique d'opinion. Celui-ci est basé sur un moteur d'analyse linguistique. Il permet de détecter finement les segments de texte porteurs d'opinions, de les extraire, et de leur attribuer une note selon la polarité qu'ils expriment. Nous présentons enfin les différentes perspectives que nous envisageons pour ce prototype.

ABSTRACT

FMO: a tool for automated opinion mining

We describe our prototype of automatic opinion mining. It is based on a linguistic analysis engine. It allows to subtly identifying the text phrases which bear some opinion, to extract them, and to give them a note according to the polarity that they express. Finally, we present the perspectives that we plan to carry out.

MOTS-CLÉS : Analyse d'opinion, e-reputation, extraction d'information.

KEYWORDS : Opinion mining, e-reputation, information extraction.

1 Contexte

Le traitement, manuel ou automatique, des opinions, est en pleine expansion¹, notamment en raison du développement des possibilités données au public d'exprimer son avis sur tous les sujets. Nous proposons un dispositif permettant une analyse linguistique beaucoup plus fine que la plupart des solutions existantes. Nous en décrivons les principes ci-dessous.

2 Présentation du prototype

Notre moteur de traitement est fondé sur la constatation que les opinions sont exprimées sous la forme d'expressions évaluatives, dont les principaux composants sont un *attribut* (un critère susceptible de faire l'objet d'un jugement d'opinion) et une *valeur* (l'opinion elle-même) (Garcia Villalba, 2012). Par exemple, dans des textes évaluant la qualité d'un hôtel, la chambre sera considérée comme un attribut, qui peut faire l'objet de différentes valeurs (« confortable », « propre », etc.). Notre système permet d'extraire l'attribut et la

¹ Cf. notamment analyses de BIA/Kelsey de 2010 (www.biakelsey.com/research-and-forecasts)

valeur correspondante ainsi que l'expression où elles apparaissent. Pour ces analyses, le système utilise une grammaire à plusieurs niveaux de règles, décrivant notamment les relations discursives entre attributs et arguments, ainsi que les composants lexico-syntaxiques sous-jacents (noms, adjectifs, mais aussi modificateurs tels que les adverbes, les négations ...). Par rapport à l'état de l'art, cette méthodologie permet d'identifier correctement des expressions évaluatives même lorsqu'elles sont complexes et ambiguës. Nous avons adapté à notre contexte industriel les principes de la plate-forme Textcoop (Saint-Dizier, 2012). Celle-ci permet de concevoir et implémenter ces règles sur la base d'un formalisme aisément adaptable quels que soient le domaine de langage ou l'application visée. Son architecture permet notamment de distinguer les règles de haut niveau des ressources linguistiques, et ainsi de paramétrer finement les éléments recherchés. Notre but est de pouvoir ainsi proposer à des clients, particuliers et professionnels, un outil de veille d'opinion détaillée, mais aussi synthétique.

Pour l'instant, notre système est adapté au traitement des opinions portées sur les domaines de l'hôtellerie et de la réservation de voyages, pour lesquels il atteint une performance de correction d'annotation comprise entre 76 et 85% (selon les paramètres calculés). Les règles et les ressources correspondant ont été codées manuellement, sur une durée de près de deux mois par domaine (pour une personne). Une version précédente, adaptée à la thématique de la politique, avait été mise en ligne l'année dernière pendant les élections présidentielles (www.polirama2012.fr).

Le prototype dont nous ferons la démonstration utilise une interface graphique qui permet de choisir le ou les textes à traiter, d'afficher leur contenu avant et après traitement (sous la forme d'annotations), ainsi que la synthèse détaillée des résultats obtenus pour un ensemble de textes.

3 Perspectives

Nous envisageons de nombreuses améliorations, à deux principaux niveaux. D'une part, ajouter de nouveaux traitements, dont la prise en compte de l'argumentation et des suggestions éventuelles, à notre connaissance inédit dans l'état de l'art et le marché. D'autre part, semi-automatiser l'acquisition de ressources linguistiques, une caractéristique primordiale pour adapter notre plateforme à un large panel de clients et de domaines.

Références

GARCIA VILLALBA M., SAINT-DIZIER P., Some Facets of Argument Mining for Opinion Analysis, *COMMA*, IOS PUBLISHING, VIENNE, SEPTEMBRE 2012.

P. SAINT-DIZIER, Processing Natural Language Arguments with the <TextCoop> Platform, *Journal of Argumentation and Computation*, vol 3-1, mars 2012.