

Analyse d'opinions multi-aspects pour la recommandation fine de restaurants

Isabelle Tellier¹

Hamid Hammouche²

Didier Cholvy²

Jean-Baptiste Tanguy²

¹ Lattice, CNRS (UMR8094), ENS Paris, Université Sorbonne Nouvelle - Paris 3

² advanced decision, 243 B Boulevard Pereire, 75017 Paris

isabelle.tellier@sorbonne-nouvelle.fr, hamid.hammouche@advanceddecision.fr

didier.cholvy@advanceddecision.fr, jean-baptiste.tanguy@advanceddecision.fr

Résumé

Dans cet article, nous présentons la tâche d'analyse d'opinions multi-aspects et son application pour la recommandation dans un cadre industriel. Les textes analysés concernent des avis en français sur des restaurants, postés sur Internet. Ces textes ont été manuellement annotés conformément à une ontologie du domaine, pour en extraire les différentes facettes de l'expérience utilisateur décrite et qualifier leur polarité. Plusieurs expériences ont ensuite été menées pour apprendre automatiquement à reproduire ce type d'annotation sur de nouveaux textes. L'utilisation de deux CRF successifs ont permis d'atteindre les meilleurs résultats.

Mots Clés

traitement automatique des langues, analyse multi-aspects, fouille d'opinion, CRF

Abstract

In this paper, we present the task of multi-aspect opinion mining and its application for recommendation in an industrial context. The texts used are comments in French about restaurants, posted on the Internet. These texts have been manually labeled according to a domain ontology, so as to extract the various aspects of the user experience and attach a polarity to them. Several Machine Learning experiments have been conducted to reproduce this annotation on new texts. The best results have been obtained by using two consecutive CRFs.

Keywords

natural language processing, multi-aspect analysis, opinion mining, CRF

1 Introduction

L'objectif final de nos travaux est de proposer un service industriel de recommandation personnalisée de produits touristiques. Ce service s'appuiera notamment sur l'analyse automatique de commentaires utilisateurs postés sur Internet concernant ces produits, en y recherchant les "expériences utilisateurs" fines qui y sont décrites. Cette dé-

marche correspond à la tâche d'*extraction automatique d'opinions multi-aspects* dans des textes. Dans un premier temps, nous nous sommes concentrés sur le domaine de la restauration. Dans ce cadre, les différents "aspects" pertinents pouvant apparaître dans un commentaire sont par exemple l'ambiance, le cadre, la cuisine, le service, etc. Chacun d'eux décrit une facette de l'expérience utilisateur et peut être qualifié positivement ou négativement.

Nous avons constitué un corpus original en français de commentaires et l'avons manuellement annoté en distinguant 12 aspects distincts et en associant une polarité (au minimum : positive, négative ou neutre) à chaque occurrence de ces aspects. Nous avons ensuite utilisé ce corpus pour procéder à des expériences d'apprentissage automatique visant à reproduire cette annotation sur de nouveaux textes. Comme cela est traditionnellement fait dans le domaine, les deux étapes d'identification des aspects et de qualification de leur polarité ont été traitées séquentiellement : la première par un CRF, la deuxième en mettant en concurrence un SVM et un nouveau CRF. Sur nos données, c'est la deuxième approche qui s'est avérée la plus performante.

Dans la suite de l'article, nous présentons tout d'abord rapidement l'état de l'art du domaine. Nous décrivons ensuite le corpus que nous avons constitué, puis les expériences que nous avons menées et leurs résultats.

2 Etat de l'art

L'analyse d'opinion multi-aspects est une sous-tâche bien identifiée du Traitement Automatique des Langues [4, 6, 5], elle a notamment fait l'objet d'une compétition SemEval en 2014 [9]¹, 2015 [8]² et 2016 [7]³. Elle est traditionnellement traitée en deux étapes distinctes : identification des "aspects" d'abord puis qualification de leur polarité. Pour l'identification des aspects, quelques approches à base de patrons linguistiques manuellement définis ont été testées, mais ce sont bien sûr les méthodes d'apprentissage automatique supervisé qui ont depuis quelques années pris

1. <http://alt.qcri.org/semeval2014/task4/>

2. <http://alt.qcri.org/semeval2015/task12/>

3. <http://alt.qcri.org/semeval2016/task5/>

le dessus. Comme un "aspect" est un empan de texte (généralement supposé contigu) pouvant couvrir plusieurs mots, ce sont les techniques d'annotation de séquences qui sont les plus adaptées, avec un codage BIO (Begin/In/Out) pour bien en délimiter les frontières. Les CRF [12], et plus récemment les réseaux de neurones convolutifs [10] ont montré leur efficacité en la matière. Associer une polarité à un aspect est une tâche de classification simple, qui a été traitée notamment par de la régression logistique [3], des SVM [1] ou des réseaux neuronaux [11].

3 Constitution du corpus annoté

Lors de SemEval 2016, une partie des données proposées étaient en français et portaient sur des restaurants [2]. Les données d'entraînement et de test de ce corpus comprennent en tout 457 textes correspondant à 2365 phrases. Notre corpus a été développé indépendamment⁴. Il contient 8381 avis issus de différentes sources. Ces avis portent sur des restaurants parisiens, ils correspondent à 46213 phrases : un avis a en moyenne 6 phrases, 75 unités (mots, émoticônes...) dont 51 distincts. Les annotations des deux corpus ne sont pas exactement les mêmes :

- le corpus de SemEval est annoté au niveau des phrases, une propriété "target" permet de citer la portion de texte correspondant à l'aspect annoté : cette propriété est parfois vide, parfois dupliquée dans les annotations. Notre corpus est annoté au niveau des unités linguistiques, chaque unité ne peut donc recevoir qu'une seule étiquette.
- les étiquettes du corpus SemEval sont des paires Entité-Attribut (6 entités, auxquelles peuvent être attachés certains attributs : 12 paires possibles en tout). Notre jeu d'étiquettes, inspiré d'ontologies du domaine, est aussi composé de 12 unités, mais elles ne coïncident que partiellement (certaines sont plus orientées "avis" qu'"aspect" proprement dit) : Ambiance, Astuce, Cadre, Boisson, Cuisine-Générale, Nourriture, Emplacement-Géographique, Public-Cible, Qualité-Prix, Service, Opinion générale et Recommandation.
- la polarité dans SemEval peut valoir "positif", "négatif" ou "neutre", nous avons ajouté les modalités "très positif" et "très négatif".

Pour toutes ces raisons, les résultats de nos expériences seront difficilement comparables avec ceux de la littérature.

Pour préparer nos expériences, des pré-traitements linguistiques ont été appliqués sur nos textes : segmentation, étiquetage POS (Treetagger pour l'oral et pour l'écrit, SEM⁵, Talismane⁶), lemmatisation (Talismane), chunking et entités nommées (SEM), analyse syntaxique (Talismane). Des lexiques spécialisés (termes de cuisine par exemple) ainsi que des lexiques d'opinion ont également été constitués.

4. il ne peut malheureusement pas être diffusé librement pour des raisons de confidentialité industrielle

5. <http://www.lattice.cnrs.fr/sites/itellier/SEM.html>

6. <http://redac.univ-tlse2.fr/applications/talismane.html>

4 Expériences

Pour ces expériences, nous avons systématiquement utilisé le protocole de la validation croisée à 5 plis.

Conformément à la littérature, nous avons d'abord cherché à retrouver les "aspects", indépendamment de leur polarité. Nous avons pour cela utilisé des CRF, tels qu'implémentés dans Wapiti⁷. Les textes ont été traités phrase par phrase, mais en gardant en indice leur position dans le texte. Les "patrons" (templates) permettant de générer les fonctions caractéristiques les plus efficaces comprenaient la prise en compte des tokens courant, précédant et suivant, d'expressions régulières pour les chiffres et les monnaies, des lexiques. Les meilleurs résultats obtenus sont mesurés en micro-moyennes des précisions (P), rappels (R) et F1-mesures (F1) à différents niveaux de granularité :

- pour chaque étiquette distincte : P=73,52, R=72,76, F1=73,13
- pour chaque "segment de texte" (commençant par un B d'un certain type, se prolongeant éventuellement par des I de même type) : P=57,59, R=54,23, F1=55,85
- par segment sans tenir compte de leur type : P=70,72, R=66,27, F1=68,41.

Pour identifier les polarités (classification parmi les 5 classes possibles), nous avons testé des SVM et, de nouveau, des CRF. Nous avons d'abord réalisé des expériences en utilisant les segments de texte "Gold". Pour les SVM⁸, les segments de textes ont été transformés en vecteurs d'indices comprenant divers nombres d'occurrences et proportions (des différentes catégories de mots lexicaux, mots appartenant aux lexiques d'opinion, ponctuations, mots en majuscules, indices de négation). Le meilleur SVM testé (noyau RBF, stratégie "un contre tous", vote pondéré par la précision du modèle associé à l'étiquette proposée) a atteint P=39,56, R=43,3, F1=41,19. Ce SVM n'a pas directement été confronté au vocabulaire des segments, ce qui explique sans doute ses faibles performances. Quant au CRF, il n'a pas à réaliser de segmentation : c'est donc en fait plutôt un modèle de "maximum d'entropie" qui a été appris. Les attributs utilisés se sont réduits aux tokens, aux lemmes, aux POS et à l'appartenance aux lexiques d'opinion. Les résultats ont alors atteint : P=R=F1=74,79, et même P=R=F1=80,83 quand on se ramène à 3 polarités au lieu des 5 initiales. Mis bout à bout, les meilleurs modèles de deux tâches ont la performance suivante (par bloc de segments typés) : P=35,18, R=35,27, F1=35,23.

5 Conclusion

Nous avons reporté dans cet article des expériences préliminaires visant à extraire des informations fines de commentaires en français. La tâche est difficile, nous l'avons évaluée de la façon la plus exigeante possible. Les CRF ont pour l'instant été privilégiés, pour leur utilisation facile de ressources linguistiques externes. Nous comptons par la suite tester des alternatives neuronales à nos modèles.

7. <https://wapiti.limsi.fr/>

8. avec la librairie Python sklearn

Références

- [1] T. Álvarez López, J. Juncal-Martínez, M. Fernández-Gavilanes, E. Costa-Montenegro, and F. J. González-Castaño. Gti at semeval-2016 task 5 : Svm and crf for aspect detection and unsupervised aspect-based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 306–311, San Diego, California, June 2016. Association for Computational Linguistics.
- [2] M. Apidianaki, X. Tannier, and C. Richart. Datasets for aspect-based sentiment analysis in french. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation LREC 2016, Portorož, Slovenia, May 23-28, 2016.*, 2016.
- [3] H. Hamdan. Sentsys at semeval-2016 task 5 : Opinion target extraction and sentiment polarity detection. In *Proceedings of the 10th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2016, San Diego, CA, USA, June 16-17, 2016*, pages 350–355, 2016.
- [4] M. Hu and B. Liu. Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, Washington, USA, August 22-25, 2004*, pages 168–177, 2004.
- [5] B. Liu. *Sentiment Analysis and Opinion Mining*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers, 2012.
- [6] B. Pang and L. Lee. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2) :1–135, 2007.
- [7] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. Al-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, O. D. Clercq, V. Hoste, M. Apidianaki, X. Tannier, N. V. Loukachevitch, E. Kotelnikov, N. Bel, S. M. J. Zafra, and G. Eryigit. Semeval-2016 task 5 : Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2016, San Diego, CA, USA, June 16-17, 2016*, pages 19–30, 2016.
- [8] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, and I. Androutsopoulos. Semeval-2015 task 12 : Aspect based sentiment analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2015, Denver, Colorado, USA, June 4-5, 2015*, pages 486–495, 2015.
- [9] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar. Semeval-2014 task 4 : Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation, SemEval@COLING 2014, Dublin, Ireland, August 23-24, 2014.*, pages 27–35, 2014.
- [10] S. Poria, E. Cambria, and A. F. Gelbukh. Aspect extraction for opinion mining with a deep convolutional neural network. *Knowl.-Based Syst.*, 108 :42–49, 2016.
- [11] D. Tang, B. Qin, and T. Liu. Aspect level sentiment classification with deep memory network. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*, pages 214–224, 2016.
- [12] Z. Toh and W. Wang. DLIREC : aspect term extraction and term polarity classification system. In *Proceedings of the 8th International Workshop on Semantic Evaluation, SemEval@COLING 2014, Dublin, Ireland, August 23-24, 2014.*, pages 235–240, 2014.