

“Sentiment Aware Map” : exploration cartographique de points d’intérêt via l’analyse de sentiments au niveau des aspects

Ioan Calapodescu¹

Caroline Brun¹

Vassilina Nikoulina¹

Salah Aït-Mokhtar¹

¹ Naver Labs Europe, 6 chemin de Maupertuis, 38240 Meylan

firstname.lastname@naverlabs.com

Video : <https://vimeo.com/naverlabseurope/review/329135227/be5edf74be>

Résumé

Cet article décrit un prototype dédié à la recherche cartographique de points d’intérêts (PI) sur la base de scores de sentiments au niveau des aspects, extraits sur des commentaires utilisateur. Ce système comprend un composant “état de l’art” d’analyse de sentiments au niveau des aspects, qui extrait les polarités des opinions utilisateur relatives aux aspects spécifiques des points d’intérêts. Les polarités sont agrégées afin d’associer des évaluations aux PI. L’utilisateur peut chercher, filtrer, réordonner et visualiser les PI sur une carte en fonction d’aspects particuliers. Des profils de recherche multiples peuvent être définis via des préférences sur ces aspects.

Mots Clef

Analyse de sentiments au niveau des aspects, recherche cartographique, visualisation et filtrage de points d’intérêts

Abstract

This paper describes a proof-of-concept system for searching points-of-interest (POIs) on a map, on the basis of Aspect-Based sentiment scores extracted automatically from free text user reviews. The system includes a state of the art Aspect-Based Sentiment Analysis component that extracts the polarities of users opinions on specific POI aspects. Polarities are aggregated to assign aspect-specific ratings at the level of POIs. Users can search, filter, rerank and visualize POIs on a map according to specific aspects. Multiple search profiles can be defined as preferences on POI aspects.

Keywords

Aspect-Based Sentiment Analysis, map search, point of interest visualization and filtering

1 Introduction

De nombreuses applications web et mobiles sont dédiées à la recherche cartographique de points d’intérêts (PI), qu’elles soient génériques (“Google Maps”, “Bing”, “Naver Maps”) ou spécifiques au tourisme (“TripAdvisor”,

“Booking.com”). De telles applications permettent de chercher des PI tels que des restaurants, musées, etc. et d’obtenir un ensemble d’informations associées : adresse, téléphone, site web, itinéraire, horaires, notes des utilisateurs et opinions concernant le PI, etc. Un ensemble d’études montrent que les notes et revues des utilisateurs ont un impact important sur le nombre de visites et la rentabilité des PI [Kim et al. 2016, Phillips et al. 2017]. Les services les plus populaires permettent de chercher et d’ordonner les PI selon les appréciations globales des utilisateurs. Cependant, il n’est pas possible de le faire sur la base d’opinions associées à des aspects précis, tels que la *qualité de la nourriture*, le *prix*, l’*ambiance* pour un restaurant ou la *localisation* ou la *propreté* pour un hôtel. En fait, la vaste majorité des PI n’ont pas de notes spécifiques associées aux aspects, même si cela est prévu dans le formulaire d’évaluation¹. Par contre, les utilisateurs expriment fréquemment leurs opinions sur les aspects des PI dans les commentaires en texte libre². Nous présentons ici un prototype qui permet à l’utilisateur de chercher, d’ordonner et de visualiser les PI sur une carte, selon les scores de polarité associés aux aspects, scores calculés automatiquement à partir des commentaires en texte libre.

2 Description du système

2.1 Données et architecture globale

Ce prototype est basé sur un sous-ensemble de données issues de la plateforme Foursquare³, qui comprend environ 100M PI. Nous avons analysé les revues (47K) de PI (15K) d’une capitale européenne, en français et en anglais. L’architecture globale est représentée sur la figure 1. L’application fournit un “frontend web”, pour les interactions de recherche par l’utilisateur, filtrage des résultats et gestion des préférences. L’interface utilisateur (UI) web permet aussi la visualisation des résultats comme la liste des PI résultants, les résumés des opinions, et l’affichage des

1. Par exemple, “TripAdvisor” définit des aspects pour les restaurants, mais seule une minorité de revues utilisateur contiennent des notes pour les aspects

2. Voir par exemple, la tâche 5 de SemEval2016 [Pontiki et al. 2016].

3. <https://foursquare.com/>

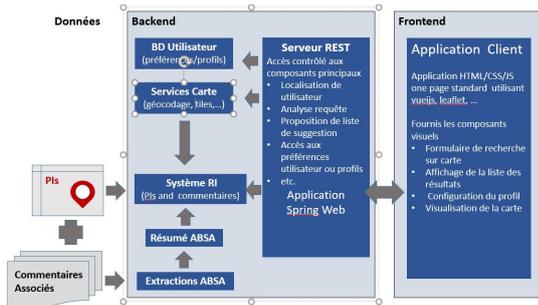


FIGURE 1 – Architecture de “Sentiment Aware Map”

aspects & des polarités sur la carte. “Frontend” et “backen-
d” interagissent via des appels à des services web de type REST. Le “backend” fournit les extractions et résumés des opinions (section 2.2 et 2.3), les services dédiés à la cartographie (géocodage et géocodage inverse) et les fonctionnalités de recherche d’information (RI). Les PI sont tout d’abord indexés dans le serveur SOLR avec leur nom, adresse, type, horaires, coordonnées géographiques, etc. Les revues associées à ces PI sont ensuite analysés par un détecteur de langue puis par le composant KET-ABSA (section 2.2) qui analyse les sentiments au niveau des aspects (ABSA). Les résultats sont résumés et stockés dans l’index RI.

Ces différents composants sont détaillés dans ce qui suit.

2.2 Le composant KET-ABSA

L’analyse de sentiments au niveau des aspects (Aspect Based Sentiment Analysis en anglais, ABSA) s’attelle à extraire et résumer les opinions décrivant l’avis des utilisateurs sur des entités spécifiques et sur leurs aspects, c’est-à-dire les différentes caractéristiques qui les qualifient. Les aspects sont des attributs de ces entités, par exemple le *service* dans un restaurant ou la *qualité de l’écran* d’un téléphone portable, généralement décrits sous forme d’une ontologie. Un système de détection des sentiments au niveau des aspects identifie l’entité sur laquelle une opinion s’exprime, qualifie la nature de l’aspect associé et lui associe la polarité correspondante (positive, négative ou neutre).

Nous utilisons le système KET-ABSA (c.f. Figure 2), que nous avons développé [Brun et al. 2016], afin d’extraire les sentiments au niveau des aspects dans les revues des PI. L’approche adoptée combine de l’information linguistique riche à des modèles d’apprentissage machine, approche récemment améliorée par l’utilisation de plongements de mots et de ressources lexicales acquises via des méthodes faiblement supervisées [Brun and Nikoulina 2018]. KET-ABSA améliore les meilleurs scores obtenus lors de SemEval16 pour la détection des aspects et des polarités, sur les revues de restaurants en anglais et en français (voir le tableau 1).

	Tâches	SemEval16 (Max.)	KET-ABSA
EN	Termes - F1	72.3	71.4
	Aspects - F1	73.0	75.1
	Polarités - Acc.	88.1	88.4
FR	Termes - F1	66.7	66.6
	Aspects - F1	61.2	69.1
	Polarités - Acc.	78.8	82.5

TABLE 1 – Performances de KET-ABSA (Restaurants)

2.3 Aggrégation des scores

Généralement, les analyses ABSA s’effectuent phrase à phrase pour capturer finement l’information sur les aspects. Le nombre de revues d’un même PI pouvant être grand, il est nécessaire d’aggréger cette information au niveau de la revue, puis du PI afin de générer une synthèse des sentiments au niveau des aspects, ce qui est particulièrement important pour la navigation sur une carte. Pour cela, nous capitalisons sur les résultats obtenus sur les phrases par KET-ABSA, dont les classifieurs fournissent la distribution des probabilités sur l’ensemble des aspects et polarités des phrases. Le score $s(a_i, r)$ associé à l’aspect a_i au niveau de la revue est calculé comme une moyenne des probabilités de a_i pour chaque phrase $\sum_{k=1}^N p(a_i, s_k)/N$. De la même façon, le score de polarité pour chaque aspect au niveau de la revue $s(p_i, r)$ ($p_i \in \{\text{positive, negative, neutral}\}$) est une moyenne pondérée sur toutes les phrases, la valeur de polarité la plus élevée étant retenue pour chaque aspect. La même méthode est ensuite appliquée au niveau du PI, pour le calcul des scores $s(a_i, poi)$ and $s(p_i, poi)$.

2.4 Recherche de PI selon les sentiments

Les extractions de KET-ABSA sont utilisées à plusieurs niveaux dans le système. Tout d’abord, elles sont utilisées, à la suite d’une requête, pour ajouter de l’information sous forme visuelle dans le descriptif du PI, comme sur la figure 3, où des pictogrammes symbolisant les aspects résumement les paires aspect/opinion les plus fréquentes. Les codes de couleurs indiquent la polarité : rouge=négative, vert=positive. Cela permet à l’utilisateur de filtrer rapidement les commentaires qui l’intéressent, par exemple ceux ayant une opinion positive sur le prix  ou l’ambiance ,

ou au contraire de rejeter ceux qui ne l’intéressent pas, par exemple ceux ayant une opinion négative sur le service . Un clic sur le pictogramme choisi permet de réordonner les résultats de requête selon l’aspect correspondant. L’utilisateur peut donc filtrer et visualiser les résultats par combinaison de couples <aspect, opinion> spécifiques.

Le système permet également de visualiser les revues utilisateurs enrichies avec les extractions ABSA (voir figure 4⁴), ce qui permet d’ordonner les commentaires selon des aspects spécifiques, par opposition à un ordonnancement

4. Les étiquettes préfixées par “user-” sont de faux identifiants.

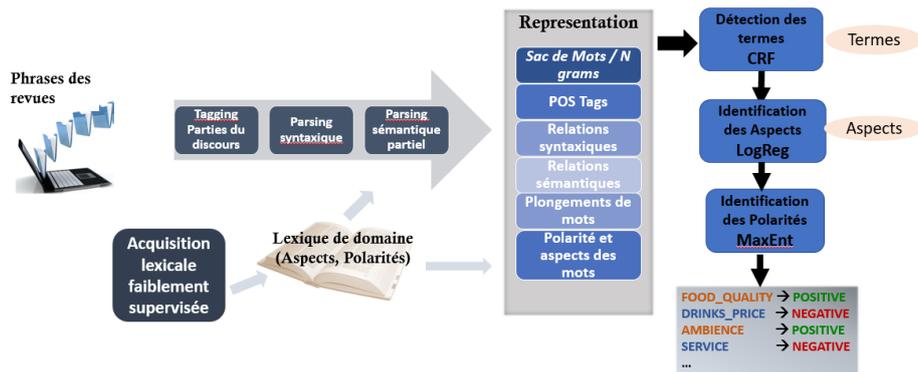


FIGURE 2 – Architecture de KET-ABSA

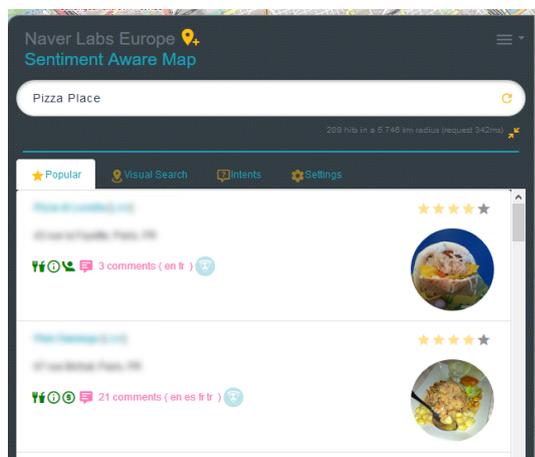


FIGURE 3 – Sentiment Aware Map : Résultats de recherche



FIGURE 4 – Revues utilisateurs enrichies par ABSA

fonction de la note globale.

Enfin, les extractions ABSA permettent de fournir un résumé en langue naturelle des revues. Par exemple, sur la figure 4, en plus de montrer que la nourriture est bonne mais pas l'ambiance, on peut montrer comment ces opinions sont exprimées dans le texte ("*Le Kastu Curry est bon*", "*Too crowded*").

2.5 Exploration et visualisation des PI

Les extractions ABSA permettent la visualisation des couples <aspect, sentiment> sur une carte. Chaque PI sur la carte est associé à un pictogramme résumant le couple le plus fréquent observé pour ce PI. L'utilisateur a trois options de visualisation : aspect positif le plus fréquent (**Top Positive**), aspect négatif le plus fréquent (**Top Negative**), ou les deux (**Mixed**). Cette visualisation donne une idée à l'utilisateur des avis concernant des régions de la carte : zones touristiques avec de fréquents avis négatifs sur le service, zones où les prix sont bas, etc.



FIGURE 5 – Sentiment Aware Map : Visualisation

2.6 Profil utilisateur personnalisé

Notre prototype permet de créer et de configurer des *Profils d'Intention* (figure 6), par exemple *Profil Professionnel*, *Profil Amis* ou *Profil Romantique*. Un profil permet d'ordonner les aspects selon leur importance dans un contexte donné, les aspects non pertinents pouvant être supprimés dans certains contextes. Un vecteur correspondant au profil d'intention associe des poids selon l'importance des aspects. Chaque PI ayant son propre vecteur défini à l'aide des extractions ABSA, les PI peuvent être réordonnés par un simple produit de ces deux vecteurs.

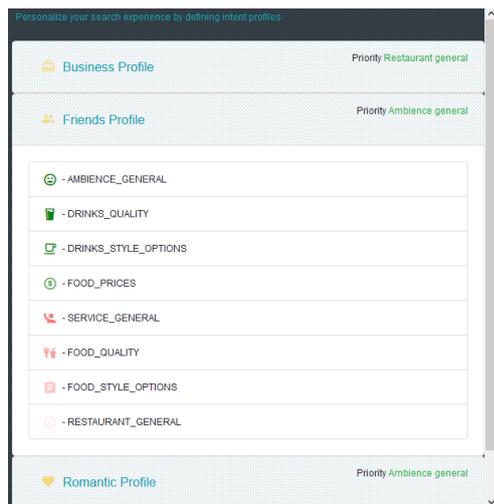


FIGURE 6 – Configuration du profil d'intention

3 Travaux connexes

[Zhao et al. 2016] décrivent *PreMiner*, un système de recommandation de PI qui capture les préférences des utilisateurs en matière de régions et d'aspects. Ce système couvre différents types de PI, les aspects les plus positifs et les plus négatifs sont visibles par région sur une carte, mais pas par PI. De plus, ce système ne permet pas de requêtes par aspect. [Baral and Li 2017] créent des profils utilisateur et des profils de PI sur la base des aspects présent dans les revues, mais les profils sont restreints aux rédacteurs de revues, et un utilisateur ne dispose que d'un seul profil. *ReviewMiner* [Wu and Wang 2017] est basé sur l'analyse latente des évaluations des aspects (LARA) [Wang et al. 2010] et prédit un score de 1 à 5 pour chaque aspect. Les aspects recherchés sont inférés à partir de la requête avec le même modèle ABSA, ce qui diffère de la notion de profil que nous proposons. Enfin, les services les plus populaires (Booking.com, Airbnb.com, Google Maps, et Tripadvisor.com) ne proposent pas de recherche selon les opinions au niveau des aspects. Certains services proposent des scores liés aux aspects, mais ils sont remplis manuellement par les utilisateurs.

4 Conclusion

Nous présentons ici un prototype pour l'exploration cartographique de points d'intérêts sur la base d'une analyse de sentiments au niveau des aspects des commentaires utilisateurs associés. Actuellement expérimenté pour le domaine restaurants, nous souhaitons l'étendre à d'autres types de PI : hotels, musées, magasins, etc.

Références

[Baral and Li 2017] Ramesh Baral and Tao Li. 2017. PERS : A personalized and explainable POI recommender system. *CoRR*, abs/1712.07727.

[Brun and Nikoulina 2018] Caroline Brun and Vassilina Nikoulina. 2018. Aspect based sentiment analysis into the wild. In *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, WASSA@EMNLP 2018, Brussels, Belgium, October 31, 2018*, pages 116–122.

[Brun et al. 2016] Caroline Brun, Julien Perez, and Claude Roux. 2016. XRCE at semeval-2016 task 5 : Feedbacked ensemble modeling on syntactico-semantic knowledge for 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 277–281.

[Kim et al. 2016] Woo Gon Kim, Jun Justin Li, and Robert A Brymer. 2016. The impact of social media reviews on restaurant performance : The moderating role of excellence certificate. *International Journal of Hospitality Management*, 55 :41–51.

[Phillips et al. 2017] Paul Phillips, Stuart Barnes, Krystin Zigan, and Roland Schegg. 2017. Understanding the impact of online reviews on hotel performance : an empirical analysis. *Journal of Travel Research*, 56(2) :235–249.

[Pontiki et al. 2016] Maria Pontiki, Dimitrios Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, Véronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia Loukachevitch, Evgeny Kotelnikov, Nuria Bel, Salud María Jiménez-Zafra, and Gülşen Eryiğit. 2016. SemEval-2016 task 5 : Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation, SemEval '16, San Diego, California*. Association for Computational Linguistics.

[Wang et al. 2010] Hongning Wang, Yue Lu, and Chengxiang Zhai. 2010. Latent aspect rating analysis on review text data : A rating regression approach. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '10*, pages 783–792. ACM.

[Wu and Wang 2017] Derek Wu and Hongning Wang. 2017. Reviewminer : An aspect-based review analytics system. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '17*, pages 1285–1288.

[Zhao et al. 2016] Kaiqi Zhao, Yiding Liu, Quan Yuan, Lisi Chen, Zhida Chen, and Gao Cong. 2016. Towards personalized maps : Mining user preferences from geo-textual data. *PVLDB*, 9(13) :1545–1548.