

---

# Fouille d'opinion : ces mots qui changent de polarité selon le domaine

**Morgane Marchand**

CEA, LIST, Laboratoire Vision et Ingénierie des Contenus  
Centre Nano-Innov Saclay, 91191 Gif-sur-Yvette Cedex<sup>1</sup>  
LIMSI-CNRS, Université Paris-Sud, 91403 Orsay Cedex<sup>2</sup>  
morgane.marchand@cea.fr

---

*RÉSUMÉ. La manière dont les gens expriment leur avis change selon ce dont ils parlent. Ainsi, un classifieur d'opinion entraîné sur des critiques de films, ne pourra pas être appliqué à des critiques de restaurants sans un minimum d'adaptation. Par exemple, certains mots présents dans les deux domaines peuvent changer de polarité. Dans cet article, nous montrons dans un premier temps que ces mots changeant de polarité affectent le comportement des classifieurs automatiques d'opinion : supprimer ces mots des corpus ou bien les considérer comme deux mots différents améliore les résultats des classifieurs. Dans un second temps, nous présentons une méthode exploratoire permettant de détecter automatiquement ces mots qui changent de polarité, sans avoir recours aux annotations du domaine cible. Cette méthode s'appuie sur la comparaison des profils de co-occurrence des mots dans les deux domaines par rapport à un ensemble de mots pivots, sélectionnés automatiquement pour n'avoir qu'une seule polarité.*

*ABSTRACT. Sentiment is expressed differently in different domains. For instance, words which are used in both domains can change their polarity. In this article, we present a study on these multi-polarity words. We first show that these multi-polarity words are actually used by automatic opinion classifiers and affect their performances: the classification performance improves if the multi-polarity words are considered as two different words, or even suppressed from the representation. In a second part, we present an exploratory method for multi-polarity words detection without any labelled target data. This method relies on the comparison of the co-occurrence profiles of the words with a set of pivot words, selected to have only one polarity.*

*MOTS-CLÉS : Fouille d'opinion, Adaptation au domaine, Changement de polarité*

*KEYWORDS: Opinion mining, Transfert learning, Multi-polarity words*

---

1. Co-directeur : Romaric Besançon, Encadrant : Olivier Mesnard
2. Directrice de thèse : Anne Vilnat

## 1. Introduction et état de l'art

L'intérêt pour la fouille d'opinion s'est développé en même temps que se sont multipliés les blogs et autres plate-formes où les internautes expriment librement leur opinion. La très grande quantité de données disponibles oblige à avoir recours à des traitements automatiques de fouille d'opinion (Pang *et al.*, 2007). Cependant, la manière dont les gens expriment leur avis change selon ce dont ils parlent (Riloff *et al.*, 2003). Les distributions des mots utilisés sont différentes d'un domaine à l'autre et un classifieur d'opinion entraîné sur des critiques de films ne pourra pas être appliqué à des critiques de restaurants sans un minimum d'adaptation. Certains mots présents dans les deux domaines peuvent de plus changer de polarité<sup>1</sup> (Navigli, 2012), et donc induire des erreurs de classification (Wilson *et al.*, 2009). Par exemple, un navet est un très mauvais film alors que pour un restaurant, parler d'une soupe de navets est parfaitement neutre. La plupart des travaux à propos de la polarité en contexte visent à améliorer un lexique général pré-existant, par exemple en donnant aux polarités des mots des poids dépendant du domaine (Choi *et al.*, 2009). Ce lexique particularisé peut ensuite être utilisé par un classifieur à base de règles. Pour les classifieurs à base de corpus, les études ont majoritairement porté sur la représentation des données (Huang *et al.*, 2012). En effet, l'erreur d'adaptation d'un classifieur dépend de sa performance sur le domaine source ainsi que la différence de distribution entre les domaines source et cible (Ben-David *et al.*, 2007). Avec une bonne projection, on peut ainsi établir des liens entre les mots du domaine cible absents du domaine source et les autres (Pan *et al.*, 2010, Blitzer *et al.*, 2007). Peu de travaux s'intéressent à ce que la prise en compte des mots changeant de polarité peut apporter à des classifieurs automatiques à base de corpus. Une exception notable est (Yoshida *et al.*, 2011), qui utilisent une formulation bayésienne du problème et se focalisent plus précisément sur l'influence du nombre de domaines source et cible, allant jusqu'à quatorze domaines différents.

Dans un premier temps, nous présentons une étude de ces mots qui changent de polarité en mettant en évidence leur influence sur les performances des classifieurs automatiques d'opinion. Dans un second temps, nous présentons une méthode destinée à détecter ces changements de sens de manière non supervisée.

## 2. Etude des mots qui changent de polarité

Nous avons utilisé le corpus anglais *Multi-Domain Sentiment Dataset*, collecté par (Blitzer *et al.*, 2007). Il s'agit de quatre corpus thématiques (*DVDs*, *kitchen*, *electronics* et *books*) contenant des critiques collectées sur le site internet Amazon. Chacun des corpus thématiques contient 1000 critiques positives et 1000 critiques négatives en plus d'un certain nombre de critiques sans étiquette. Les critiques sont représentées en sac de mots en utilisant les unigrammes et les bigrammes présents. Dans la suite de cet article nous parlerons de "mots" pour désigner indifféremment les unigrammes et les bigrammes.

---

1. La polarité est ici considérée comme le placement sur l'axe continu positif/négatif.

### 2.1. Sélection supervisée des mots changeant de polarité.

Pour chaque paire de corpus thématique, nous utilisons les critiques étiquetées. Pour chaque mot, nous regardons si sa distribution dans les critiques positives et négatives est statistiquement différente selon les domaines<sup>2</sup>. Pour ce faire, nous utilisons un test du  $\chi^2$  avec un risque de première espèce (i.e. risque de faux positif) de 1 %.

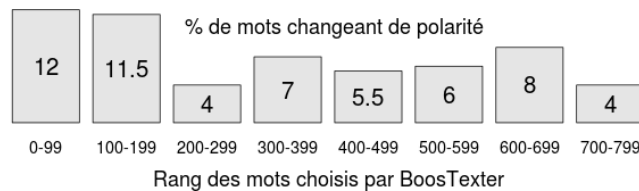
Mot	region	I loved	worry	compare	return
electronics	0.154	0.091	0.929	0.846	0.055
books	0.818	0.735	0.3	0.263	0.633

**Tableau 1.** Pourcentage de présence dans les critiques positives pour deux domaines. Le score va de 0 (très fortement négatif) à 1 (très fortement positif). Un écart de 0.5 est donc très significatif, faisant passer un mot de neutre à fortement valué.

Le tableau 1 présente quelques mots détectés comme changeant de sens. Les raisons des changements de polarité sont multiples. Par exemple, la phrase "I loved this book" est positive alors que "I loved this camera" signifie en général que maintenant ce n'est plus le cas pour cause de panne. "I loved" prend donc une valeur négative dans le corpus *electronics*, ce qui n'est pas le cas du présent "I love" qui lui reste positif.

### 2.2. Influence des mots changeant de polarité sur les classifieurs

Dans un second temps, nous avons vérifié que ces mots influencent le fonctionnement des classifieurs automatiques d'opinion.



**Figure 1.** Nombre de mots changeant de polarité parmi les mots sélectionnés par boostexter par tranche de 100 mots

Nous avons tout d'abord utilisé une méthode de boosting, BoosTexter (Schapire *et al.*, 2000), car il est possible de voir quels mots sont choisis comme classifieurs faibles. Plus un mot est sélectionné tôt, plus il est jugé pertinent pour la tâche de classification. Pour chaque paire source-cible, nous nous sommes intéressés aux rangs des mots

2. Certains mots peuvent changer de polarité à l'intérieur du même domaine mais nous ne nous intéressons ici qu'à la polarité au niveau global.

changeant de polarité et retenus par BoosTexter comme classifieurs faibles pertinents. La figure 1 présente le nombre moyen de mots changeant de polarité par tranche de 100 classifieurs faibles. Parmi les 100 premiers classifieurs faibles, ainsi que les 100 suivants, 12 % changent de polarité. Nous avons également vérifié l'influence de ces mots sur un autre type de classifieurs usuels, les SVM (SVM-light, avec un noyau linéaire), en effectuant deux expériences : les mots changeant de polarité sont soit supprimés des corpus soit différenciés (mot-SOURCE, mot-CIBLE).

	D-K	E-D	K-B	B-E	D-E	B-K
Normal	76.80	72.27	70.21	72.82	73.55	75.76
En distinguant les mots	77.58	74.04	71.01	73.90	74.01	76.38
En enlevant les mots	78.05	73.72	71.08	73.73	74.27	76.46

**Tableau 2.** Résultat d'accuracy pour un classifieur entraîné sur un domaine source et testé sur un domaine cible (S-C); D : DVD, B : books, E : electronics, K : kitchen.

Le tableau 2 montre que cette prise en compte différenciée élémentaire des mots changeant de polarité améliore la classification<sup>3</sup>. Cette tendance se retrouve pour tous les couples Source-Cible bien que l'amélioration observée ne soit pas toujours significative. Il est de plus vraisemblable qu'une pondération des mots changeant de polarité, plutôt qu'une suppression complète, donne de meilleurs résultats (Choi *et al.*, 2009). Ces résultats justifient donc la nécessité d'une détection automatique de ces mots.

### 3. Détection automatique

Les mots étudiés dans la première partie ont été sélectionnés en s'appuyant sur des annotations des critiques du domaine source et du domaine cible. Or nous nous intéressons au cas réel fréquent où l'on dispose d'un corpus du domaine cible mais sans aucune annotation. Cette section présente donc notre méthode exploratoire pour détecter les changements de polarité sans annotation dans le domaine cible.

#### 3.1. Méthode

Notre méthode utilise une liste de mots pivots sélectionnés automatiquement de manière à ce qu'ils ne changent pas de polarité. Cette sélection est présentée dans la section suivante. Ces mots pivots, présents dans les deux domaines, servent à comparer la distribution des autres mots dans deux domaines différents. Pour chaque mot, et pour chaque domaine, on réalise son profil de co-occurrence par rapport à la liste des mots pivots. Un test du  $\chi^2$  détermine ensuite si, pour un mot donné, ses profils de co-occurrence dans le domaine source et dans le domaine cible sont statistiquement différents (le mot change de polarité) ou pas (le mot ne change pas de polarité).

3. Ne sont présentés ici que les couples pour lesquels l'amélioration est la plus significative.

### 3.2. Sélection des mots pivots

Les mots pivots sont sélectionnés en deux temps : une présélection est d'abord effectuée pour que les mots choisis ne soient pas trop représentatifs d'un domaine et qu'ils soient utiles à la classification de l'opinion pour le domaine source, ensuite un processus itératif permet d'épurer cette liste des mots pouvant changer de polarité. Tout d'abord, nous calculons l'information mutuelle entre la présence et l'absence d'un mot dans une critique et l'appartenance au domaine source ou cible. Cette information mutuelle doit être faible afin que ce mot ne soit pas représentatif d'un domaine. En utilisant les étiquettes du domaine source, nous calculons également l'information mutuelle entre la présence d'un mot dans une critique source et l'étiquette positive ou négative de celle-ci. Cette fois-ci, l'information mutuelle doit être élevée, afin que ces mots soient utiles pour la détection de la polarité. Nous avons réalisé deux types de présélections, en donnant la priorité à l'une ou l'autre de ces informations mutuelles. Une fois les mots pivots sélectionnés, nous effectuons la procédure de détection des mots changeant de polarité sur les mots pivots eux-même. Nous éliminons ainsi de la liste le mot le plus susceptible de changer de polarité. Puis nous recommençons jusqu'à ce que plus aucun mot de la liste ne soit considéré comme changeant de polarité.

### 3.3. Résultats et pistes d'amélioration

Nous avons ensuite comparé les mots obtenus automatiquement par notre méthode avec ceux obtenus en utilisant les étiquettes source et cible, comme décrit dans la première partie. Nous testons les résultats pour deux jeux de mots pivots, présélectionnés soit en privilégiant l'information mutuelle source-cible, soit l'information mutuelle positif-négatif. Comme le montre le tableau 3, la méthode automatique sélectionne beaucoup plus de mots que la méthode supervisée (433 mots). Aussi la précision est faible. On peut l'augmenter sans baisser le rappel en ne prenant que les mots pour lesquels on est le plus sûr du changement de sens (précision max).

	nb mots	précision	rappel	précision max
S-C	1828 mots	14.50 %	61.20 %	16.11 %
P-N	1604 mots	16.52 %	61.20 %	18.39 %

**Tableau 3.** Comparaison des mots sélectionnés par la méthode automatique avec ceux sélectionnés par la méthode supervisée selon deux jeux de mots pivots, présélectionnés en privilégiant soit l'information mutuelle source-cible (S-C), soit l'information mutuelle positif-négatif (P-N). Domaine source : DVD ; domaine cible : kitchen.

Les mots pertinents sont donc sélectionnés en premier par notre méthode, ce qui est très encourageant. Dans des travaux futurs, nous étudierons les seuils de sélection optimaux. On remarque également que la méthode de présélection des mots pivots joue un rôle essentiel dans les performances. Aussi allons-nous porter une attention toute particulière aux critères de sélection. De plus, nous avons pour l'instant réalisé

Morgane Marchand

nos travaux sur des représentations type sac de mots au niveau du document. Nous espérons améliorer nos résultats en utilisant à la place des fenêtres glissantes afin de calculer les co-occurrences des mots.

#### 4. Conclusion

Dans cet article, nous nous sommes intéressés au problème des mots changeant de polarité d'un domaine à un autre. Notre étude a montré que pour au moins deux types de classifieurs automatiques usuels, ces mots sont effectivement utilisés. Les prendre en compte améliore les performances de la classification de l'opinion lors du transfert d'un domaine à l'autre. Nous avons également présenté une méthode exploratoire pour reconnaître automatiquement ces mots sans utiliser aucune étiquette de donnée du domaine cible. Les premiers résultats sont encourageants et mettent notamment en lumière l'importance de la pré-sélection des mots pivots sur le résultat final.

#### 5. Bibliographie

- Ben-David S., Blitzer J., Crammer K., Pereira F., « Analysis of representations for domain adaptation », *Advances in neural information processing systems*, vol. 19, p. 137, 2007.
- Blitzer J., Dredze M., Pereira F., « Biographies, bollywood, boom-boxes and blenders : Domain adaptation for sentiment classification », *ACL*, 2007.
- Choi Y., Cardie C., « Adapting a polarity lexicon using integer linear programming for domain-specific sentiment classification », *EMNLP*, 2009.
- Huang F., Yates A., « Biased Representation Learning for Domain Adaptation », *EMNLP*, Association for Computational Linguistics, Jeju Island, Korea, p. 1313-1323, July, 2012.
- Navigli R., « A quick tour of word sense disambiguation, induction and related approaches », *SOFSEM 2012 : Theory and Practice of Computer Science*, p. 115-129, 2012.
- Pan S., Ni X., Sun J., Yang Q., Chen Z., « Cross-domain sentiment classification via spectral feature alignment », *WWW*, ACM, p. 751-760, 2010.
- Pang B., Lee L., « Opinion Mining and Sentiment Analysis », *Foundations and Trends in Information Retrieval*, vol. 2, p. 1-2, 2007.
- Riloff E., Wiebe J., « Learning Extraction Patterns for Subjective Expressions », *EMNLP*, p. 105-112, 2003.
- Schapire R., Singer Y., « BoosTexter : A boosting-based system for text categorization », *Machine learning*, vol. 39, n° 2, p. 135-168, 2000.
- Wilson T., Wiebe J., Hoffmann P., « Recognizing Contextual Polarity : An Exploration of Features for Phrase-Level Sentiment Analysis », *Computational Linguistics*, 2009.
- Yoshida Y., Hirao T., Iwata T., Nagata M., Matsumoto Y., « Transfer Learning for Multiple-Domain Sentiment Analysis - Identifying Domain Dependent/Independent Word Polarities », *AAAI*, 2011.