

Étude de différentes stratégies d'adaptation à un nouveau domaine en fouille d'opinion

Anne Garcia-Fernandez Olivier Ferret
CEA, LIST, Laboratoire Vision et Ingénierie des Contenus
Gif-sur-Yvette, F-91191 France.
`prénom.nom@cea.fr`

RÉSUMÉ

Le travail présenté dans cet article se situe dans le contexte de la fouille d'opinion et se focalise sur la détermination de la polarité d'un texte en adoptant une approche par apprentissage. Dans ce cadre, son objet est d'étudier différentes stratégies d'adaptation à un nouveau domaine dans le cas de figure fréquent où des données d'entraînement n'existent que pour un ou plusieurs domaines différents du domaine cible. Cette étude montre en particulier que l'utilisation d'une forme d'auto-apprentissage par laquelle un classifieur annote un corpus du domaine cible et modifie son corpus d'entraînement en y incorporant les textes classés avec la plus grande confiance se révèle comme la stratégie la plus performante et la plus stable pour les différents domaines testés. Cette stratégie s'avère même supérieure dans un nombre significatif de cas à la méthode proposée par (Blitzer *et al.*, 2007) sur les mêmes jeux de test tout en étant plus simple.

ABSTRACT

Study of various strategies for adapting an opinion classifier to a new domain

The work presented in this article takes place in the field of opinion mining and aims more particularly at finding the polarity of a text by relying on machine learning methods. In this context, it focuses on studying various strategies for adapting a statistical classifier to a new domain when training data only exist for one or several other domains. This study shows more precisely that a self-training procedure consisting in enlarging the initial training corpus with texts from the target domain that were reliably classified by the classifier is the most successful and stable strategy for the tested domains. Moreover, this strategy gets better results in most cases than (Blitzer *et al.*, 2007)'s method on the same evaluation corpus while it is more simple.

MOTS-CLÉS : fouille d'opinion, adaptation à un nouveau domaine, auto-apprentissage.

KEYWORDS: opinion mining, domain adaptation, self-training.

1 Introduction

Le travail présenté dans cet article part de deux constats bien connus en Traitement Automatique des Langues, et notamment en fouille d'opinion, lorsqu'une approche par apprentissage supervisé est mise en place. D'une part, la constitution de ressources annotées est un processus très coûteux. D'autre part, un système ayant de bonnes performances dans un domaine donné n'est pas nécessairement adapté à un autre domaine. De nombreux travaux ont été menés en

fouille d'opinion, en particulier sur la problématique de l'adaptation à un nouveau domaine. Les approches ainsi développées vont de l'identification des correspondances terminologiques entre domaines à l'utilisation de ressources externes telles que des lexiques d'opinion, lexiques au sein desquels chaque terme se voit typiquement associer une *polarité* positive, négative voire même neutre. Mais que se passe-t-il si l'on ne dispose pas de telles ressources ? Comment procéder en n'ayant accès qu'à quelques données annotées ne relevant pas du domaine ciblé ?

Nous nous attachons dans cet article à étudier des stratégies d'apprentissage automatique pour la classification de textes n'utilisant pas de ressources externes. Différentes configurations de données annotées pour la tâche de classification de textes en polarité positive ou négative sont testées afin de déterminer celles permettant d'obtenir les meilleures performances lorsqu'il s'agit de classer des textes relevant d'un nouveau domaine. En particulier, nous exposons une approche d'apprentissage dit *itératif* par laquelle un corpus annoté manuellement est enrichi au cours de boucles successives par des données annotées automatiquement.

2 Fouille d'opinion et classification de textes multi-domaine

La fouille d'opinion regroupe un grand nombre de travaux centrés sur l'exploration de textes afin d'en déterminer le caractère objectif ou subjectif ou encore d'extraire les avis qui y sont exprimés par leurs auteurs. Une des caractéristiques de ce type d'analyses est leur caractère transversal : leur nature ne dépend pas le plus souvent du domaine des textes considérés. Ainsi, dans le travail considéré ici, qui se focalise sur une tâche de classification de textes subjectifs, en l'occurrence des critiques issues du site AMAZON, les deux classes considérées, polarité positive ou négative, sont générales et non dépendantes des différents domaines qu'abordent ces critiques. En revanche, les moyens pour effectuer cette classification peuvent être plus ou moins liés à un domaine, ce qui rend la problématique de l'adaptation à un nouveau domaine de ce type d'analyses particulièrement importante. Dans ce qui suit, nous illustrerons d'abord le caractère contextuel de la détermination de la polarité d'un énoncé avant de passer en revue les principales approches pour l'adaptation d'une telle classification à un nouveau domaine.

2.1 Classer des textes selon leur polarité

Une des principales approches pour déterminer la polarité d'un texte consiste à se focaliser sur des termes porteurs d'opinion. De tels termes peuvent être trouvés dans des ressources de référence, telles que SENTIWORDNET (Esuli et Sebastiani, 2006). Néanmoins, la disponibilité de ces ressources n'est pas suffisante pour déterminer la polarité d'un texte. En effet, la polarité d'un terme peut dépendre de son contexte. Cette dépendance existe d'abord à un niveau local. Ainsi, la présence d'une négation dans une phrase peut inverser la polarité de celle-ci alors même qu'elle contient un ou plusieurs termes négatifs ("Cela dit le réalisateur sait y faire et c'est bien pour ça que le film n'est pas mauvais du tout, ni ennuyeux, ni lent."). Cette dimension est prise en compte par de nombreux travaux. Taboada *et al.* (2011) utilisent ainsi un lexique de termes porteurs d'opinion mais pondèrent l'importance de ces termes en fonction du caractère subjectif ou objectif des paragraphes dans lesquels ils apparaissent. Ils intègrent par ailleurs les négations, les modificateurs (notamment les modificateurs d'intensité) et la modalité (en particulier les suppositions). Choi et Cardie (2009) proposent au travers de l'algorithme Vote & Flip de

prendre en compte la présence d'une ou plusieurs négations dans le contexte d'un terme porteur d'opinion modifiant ainsi sa polarité. L'objet auquel un terme porteur d'opinion se rapporte peut également influencer sur la valeur de ce terme. Ainsi *mortel* est porteur d'opinion négative dans *un ennui mortel* mais d'opinion positive dans *cette fête était vraiment mortelle*. Enfin, la polysémie des termes peut aussi jouer un rôle puisque certains termes n'ont pas la même valeur d'opinion selon leur sens. C'est le cas par exemple de *navet* qui, dans la phrase "C'est un navet.", n'est pas porteur d'opinion si l'on fait référence au légume mais renvoie à une opinion négative s'il qualifie un film. La polarité d'un terme dans un texte est alors influencée plus globalement par le domaine auquel ce texte se rattache. Dans cette perspective, Harb *et al.* (2008) proposent d'acquérir un lexique d'opinion lié à une thématique en sélectionnant dans des corpus liés à cette thématique des termes cooccurents avec des termes dont la polarité est déjà connue et en utilisant une mesure de similarité entre les termes candidats et les termes connus.

2.2 D'un domaine à l'autre

Dans le prolongement des travaux de la section précédente fondés sur des lexiques d'opinion, une première voie pour aborder le problème de la dépendance par rapport au domaine en matière de fouille d'opinion consiste à définir des capacités d'adaptation automatique de ces lexiques à un domaine donné. Dans cette optique, Jijkoun *et al.* (2010) proposent d'adapter un lexique d'opinion général à un domaine spécifique en caractérisant les termes de ce lexique par des profils de contextes syntaxiques obtenus à partir d'un corpus général. Ces profils sont ensuite utilisés pour identifier les termes porteurs d'opinion pertinents pour ce domaine à partir de corpus représentatifs de celui-ci. Gindl *et al.* (2010) adaptent pour leur part un lexique d'opinion en fonction du domaine considéré en supprimant les termes de ce lexique dont la polarité varie selon le contexte.

Outre l'utilisation de lexiques d'opinion, la détermination de la polarité d'un texte peut bénéficier de l'utilisation de corpus annotés. La dépendance de ceux-ci par rapport à un domaine donné rend néanmoins leur usage délicat et conduit à définir différentes approches pour compenser cette dépendance. Denecke (2009) associent ainsi le lexique général SENTIWORDNET et un corpus d'entraînement relevant de différents domaines autres que le domaine cible pour classer des textes comme subjectifs ou objectifs. Pour la même tâche, Aue et Gamon (2005) s'affranchissent de lexiques d'opinion et comparent différentes approches utilisant des corpus d'apprentissage d'un domaine autre que le domaine cible. La disponibilité de corpus annotés relevant de différents domaines est ainsi un élément clef dans la tâche de classification de textes d'un nouveau domaine. Blitzer *et al.* (2007) construisent un tel corpus, le *Multi-domain Sentiment Dataset* (MDSD), en s'appuyant pour minimiser les coûts d'annotation sur les critiques rédigées sur le site AMAZON portant sur des objets variés appartenant à 25 grands domaines. Par ailleurs, Blitzer *et al.* (2007) proposent également de chercher des correspondances entre domaines par la technique du *Structural Correspondence Learning* (SCL) en s'appuyant sur des traits et des prédicteurs pivots. Li et Zong (2008) réutilisent ce corpus et proposent deux approches. L'approche par fusion de traits (ou *feature fusion*) se fonde sur le regroupement des corpus d'apprentissage de différents domaines en un seul. L'approche par fusion de classifieurs (ou *classifier fusion*) consiste à construire autant de classifieurs que de domaines sources disponibles et à entraîner un classifieur (un méta-classifieur) sur les sorties de ces modèles.

Quelle que soit la méthode, les performances obtenues pour des textes relevant d'un nouveau domaine cible sont moins bonnes que celles obtenues en disposant de données annotées dans le

domaine cible. Pour réduire le coût de cette annotation, des approches dites d'*Active learning* ont été proposées. L'idée est de détecter les exemples classés avec un faible score de confiance par le modèle, d'annoter ces exemples manuellement et d'intégrer ces nouvelles données aux données d'apprentissage. Si ces méthodes permettent de limiter la quantité d'annotation manuelle à effectuer, elles restent tout de même coûteuses. De ce fait, des méthodes dites d'auto-apprentissage (*self-training*) proposent, à l'image de (Drury *et al.*, 2011), de s'appuyer sur des données non annotées manuellement mais classées avec un fort score de confiance par un classifieur pour enrichir le corpus d'apprentissage de ce dernier et élargir ainsi sa couverture. C'est l'approche que nous privilégierons ici en la transposant au cas de textes appartenant à d'autres domaines.

3 Stratégies d'adaptation

L'étude que nous présentons dans cet article aborde la classification de textes en termes de polarité positive ou négative selon une approche supervisée fondée sur un corpus d'entraînement annoté manuellement, sans s'appuyer sur un lexique d'opinion constitué *a priori*. Dans ce cadre, qui reprend celui de (Blitzer *et al.*, 2007) et de (Li et Zong, 2008), son objectif est de déterminer, partant de corpus d'entraînement dans un ou plusieurs domaines sources et d'un corpus non annoté dans un domaine cible (corpus dit de développement), la stratégie la plus adaptée de constitution d'un nouveau corpus d'entraînement à partir de ces corpus disponibles afin d'obtenir les meilleures performances possibles sur le domaine cible. Les différentes stratégies considérées se différencient selon deux facteurs principaux : l'utilisation de corpus appartenant à un seul ou à plusieurs domaines sources ; l'utilisation ou non d'un corpus du domaine cible non annoté. Nous avons plus précisément testé les stratégies suivantes, chacune reposant sur le même volume de textes annotés manuellement pour constituer leur corpus d'entraînement :

Un corpus source [BASELINE] Cette stratégie baseline utilise un corpus annoté d'un unique domaine source autre que le domaine cible.

Apprentissage itératif à partir d'un corpus source [ITE-FIXE, ITE-SEUIL] Dans cette approche, nous entraînons un modèle sur un seul corpus source, comme précédemment, classifions les textes du corpus du domaine cible, sélectionnons les exemples ayant le meilleur score de confiance et intégrons ces exemples au corpus d'entraînement. Cette boucle est répétée jusqu'à épuisement du corpus du domaine cible. La sélection des meilleurs exemples peut se faire en fonction d'un seuil sur le score de confiance (approche dite ITE-SEUIL) ou bien en fonction d'un nombre fixe d'exemples intégrés à chaque itération (ITE-FIXE).

Plusieurs corpus sources [MULTI-DOMAINE] Dans cette configuration, les corpus de plusieurs domaines sources sont associés pour construire le corpus d'entraînement. Nous prenons ici tous les domaines sources, ce qui représente une autre forme de baseline.

Méthode par vote [MULTI-VOTE] Dans cette stratégie, elle aussi classique, nous entraînons un modèle par corpus source et procédons à une classification finale par vote : un exemple donné est ainsi classé en fonction de la décision majoritaire observée parmi les classifieurs associés à chaque domaine source.

Apprentissage itératif à partir de plusieurs corpus sources [ITE-MULTI-VOTE] Cette stratégie est une hybridation de la méthode par vote et de l'apprentissage itératif. À partir d'un corpus source, nous entraînons plusieurs modèles (un par domaine source) ; puis nous sélectionnons les meilleurs exemples qui sont ensuite intégrés dans chacun des corpus sources de départ. Nous sélectionnons alors les exemples classés unanimement par tous les modèles.

4 Mise en œuvre et résultats

Nos expériences ont été menées sur le corpus MDSD évoqué précédemment, corpus composé de critiques portant sur des produits variés et triées par domaine (livre, cuisine et articles ménagers, vêtements ...). Une critique est composée d'un texte, d'un titre et d'une note. Les textes sont courts, quelques phrases seulement, et rédigés en anglais. Les notes varient de 1 à 5, 1 indiquant l'avis le plus négatif sur le produit et 5 l'avis le plus positif. L'exemple présenté à la figure 1 est une critique issue du sous-corpus "Book". Nous avons utilisé la même configuration de données

```
<review>
  (...)
  <rating>5.0</rating>
  <review_text>
    I read Les Misérables after I saw the opera, and it has inspired
    in me more than any book I've ever read. I don't believe one
    could ever find a better novel anywhere. For everyone (...)
  </review_text>
</review>
```

FIG. 1 – Exemple de critique issue du MDSD relevant du domaine BOO

que celle de (Blitzer *et al.*, 2007) afin de pouvoir comparer nos résultats aux leurs à la différence près que les corpus de test ont été scindés en corpus de développement et corpus de test. Le tableau 1 présente une description générale du corpus utilisé, organisé en quatre domaines. Les données d'apprentissage sont équilibrées entre critiques positives et négatives¹ et sont présentes en même quantité dans les quatre domaines.

Domaine	Corpus d'entraînement		Corpus de développement	Corpus de test
	#critiques	#formes/critiques	# critiques	# critiques
Cuisine & articles ménagers (KIT)	2 000	96	2 000	3 945
Livres (BOO)	2 000	174	2 000	2 465
DVD (DVD)	2 000	189	2 000	3 945
Matériel électronique (ELE)	2 000	113	2 000	3 945

TAB. 1 – Description générale du corpus

À l'instar de (Torres-Moreno *et al.*, 2007), nous avons utilisé un modèle de classification à base de boosting pour effectuer nos tests, modèle mise en œuvre grâce à l'outil BoosTexter (Shapire et Singer, 2000). Le modèle produit est composé d'un ensemble de règles binaires (ou *weak learners*) portant chacune sur la présence d'un n-gramme et se voyant associer une probabilité par rapport à chaque classe considérée. Lors de la phase de classification, un score est calculé pour chaque classe en fonction des règles déclenchées par le texte traité et la classe de plus haut score est attribuée au texte. La configuration utilisée a été sélectionnée empiriquement en optimisant les paramètres pour notre approche *baseline*. Le nombre de tours a ainsi été fixé à 50. Les règles utilisent des n-grammes de taille 1 et les textes ne sont pas lemmatisés.

Les résultats de classification sont donnés en termes d'exactitude (*accuracy*) afin de pouvoir comparer nos résultats avec ceux de (Blitzer *et al.*, 2007). Le tableau 2 présente les résultats pour

¹Comme Blitzer, nous considérons une critique comme positive si sa note est > 3 et négative si elle est < 3.

l'ensemble des approches. Si l'on se concentre en premier lieu sur les stratégies ne mettant en jeu qu'un seul domaine source, on peut observer que l'approche **baseline** donne des exactitudes supérieures à 70% quels que soient les domaines d'entraînement (TRN) et les domaines de test (TEST). L'approche **ite-fixe** est simple mais ne prend pas en compte le score de confiance donné par BoosTexter. Elle ne dispose donc pas de critère naturel d'arrêt de prise en compte de nouveaux exemples, ce qui la conduit à « consommer » tout le corpus de développement. Elle obtient en pratique des performances très inférieures à celles de **baseline** et se révèle être la moins bonne de nos stratégies. La prise en compte du score de confiance de BoosTexter (approche **ite-seuil**) donne en revanche la meilleure exactitude (notée par *) pour la plupart des couples (domaine source, domaine cible). Il est à noter que les cas où cette approche ne donne pas de meilleurs résultats que **baseline** font tous intervenir les données du domaine ELE (matériel électronique).

TRN	TEST	baseline	ite-fixe	ite-seuil	multi-domaine	multi-vote	ite-multi-vote
DVD	BOO	79,7	48,8	84,4*	68,1	69,6	75,6†
ELE	BOO	75,4	41,6	79,3*			
KIT	BOO	70,9	38,1	81,8*			
BOO	DVD	77,2	69,5	82,0*	72,2	70,3	81,4†
ELE	DVD	76,2	54,3	73,4			
KIT	DVD	76,9	54,0	78,2*			
BOO	ELE	77,5*	64,5	65,5	69,1	64,7	77,5†
DVD	ELE	74,1*	69,7	62,2			
KIT	ELE	86,8*	60,4	75,7			
BOO	KIT	78,9	68,4	82,7*	75,4	71,0	81,4†
DVD	KIT	81,4	61,1	82,3*			
ELE	KIT	85,9	65,6	78,7			

TAB. 2 – Résultats en termes d'exactitude pour l'ensemble des approches

Les résultats des stratégies utilisant des données annotées relevant de plusieurs domaines sont présentés dans les 3 dernières colonnes du tableau 2. L'approche **multi-domaine** constitue dans ce cas notre **baseline**. La comparaison de ses résultats avec ceux de l'approche **baseline** (un unique domaine source) est clairement en défaveur de l'approche **multi-domaine** dans tous les cas de figure. Ce constat tend à montrer qu'utiliser un corpus d'entraînement composé de données sources hétérogènes du point de vue thématique (approche **multi-domaine**) est une moins bonne option pour classer des textes dans un domaine cible qu'utiliser un corpus d'entraînement source thématiquement homogène. Il est néanmoins possible qu'une telle observation soit à nuancer en fonction de la taille des corpus et du nombre de domaines. L'approche par vote (**multi-vote**) ne permet pas quant à elle d'obtenir une performance de classification plus élevée que la **baseline**. Cette approche, tout comme l'approche **ite-fixe**, ne prend pas en compte le score de confiance accordé par le modèle lors de la classification et le fait que la majorité des modèles catégorisent un exemple dans une classe donnée n'est apparemment pas un indice suffisant pour compenser cette insuffisance. Cependant, là encore, le nombre de domaines considérés peut avoir une influence. L'approche **ite-multi-vote** est celle des trois approches utilisant plusieurs corpus annotés donnant les meilleurs résultats (notés par †) et ce, quel que soit le domaine cible. La méthode d'apprentissage itérative se révèle donc particulièrement intéressante dans ce cas de figure comme elle l'est dans le cas d'un domaine source unique. L'exactitude moyenne de **ite-multi-vote**, égale à 79,0, est même légèrement supérieure à l'exactitude moyenne de **ite-seuil**, égale à 77,2, en particulier du fait d'un

meilleur comportement pour le domaine cible ELE.

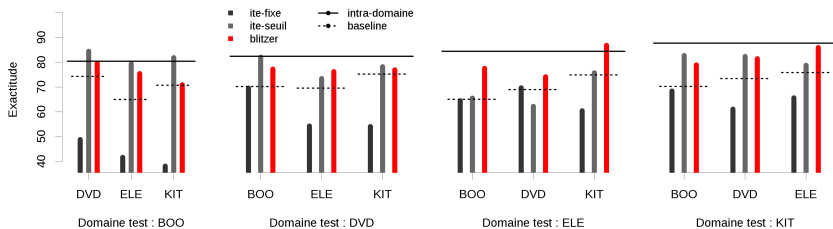


FIG. 2 – Exactitude obtenue par les approches utilisant un seul domaine d'apprentissage²

La figure 2 propose une comparaison de nos résultats avec ceux de (Blitzer *et al.*, 2007). Les barres horizontales pleines indiquent l'exactitude obtenue par une approche intra-domaine (les corpus source et cible relèvent du même domaine). Les barres horizontales pointillées indiquent l'exactitude obtenue par notre approche *baseline*. Pour chaque couple de domaines sources et cibles sont indiqués nos résultats ainsi que ceux de Blitzer. On peut observer que notre meilleure approche (*ite-seuil*) obtient de meilleurs résultats que ceux de Blitzer en dehors du domaine ELE (qu'il soit source ou cible). L'approche proposée dans (Blitzer *et al.*, 2007) permet de mettre en correspondance des termes supposés équivalents d'un domaine à un autre, termes prenant la forme de n-grammes tels que *must read* (BOO) ou *excellent product* (KIT). Il semble néanmoins que ces correspondances concernent essentiellement des termes constitués de mots pleins. Nous expliquons la performance de notre approche par le fait que nos modèles ne favorisent pas un type d'unités plutôt qu'un autre, ce qui leur permet d'utiliser aussi bien des mots outils, qui se retrouvent dans tous les domaines, que des mots pleins, plus spécifiques à un domaine. Or les mots outils jouent un rôle dans l'expression des opinions puisqu'ils permettent notamment d'exprimer la négation et l'intensité (Taboada *et al.*, 2011).

5 Conclusion et perspectives

Dans cet article, nous avons présenté une étude sur différentes stratégies possibles pour construire un classifieur statistique pour un domaine cible en ne disposant de données annotées pour son entraînement que pour un ou plusieurs autres domaines. Nous avons en particulier montré l'efficacité pour cette tâche d'une stratégie d'apprentissage itératif assimilable à une forme d'auto-apprentissage et consistant à incorporer progressivement dans le corpus d'entraînement du classifieur les textes d'un corpus du domaine cible que ce classifieur annote avec la plus grande confiance. Cette stratégie se révèle même dans un nombre significatif de cas plus efficace que la méthode présentée dans (Blitzer *et al.*, 2007) tout en étant plus simple. Une des prolongations les plus immédiates de ce travail est sa généralisation à d'autres types de classifieurs que le

²L'approche *intra-domaine* correspond à la même configuration que notre approche *baseline* à la différence que les corpus d'entraînement et de test relèvent du même domaine.

boosting utilisé ici. Au-delà, nous envisageons la transposition à notre contexte inter-domaine de la démarche d'auto-apprentissage présentée dans (Wiebe et Riloff, 2005), démarche fondée sur l'utilisation d'un classifieur supplémentaire, de nature différente du classifieur initial, pour la constitution non supervisée du corpus d'entraînement.

Remerciements

Ce travail a été financé par la Fondation Jean-Luc Lagardère. Nous tenons également à remercier Morgane Marchand et Romaric Besançon pour leur contribution aux prémices de ce travail.

Références

- AUE, A. et GAMON, M. (2005). Customizing sentiment classifiers to new domains : a case study. *In RANLP 2005*.
- BLITZER, J., DREDZE, M. et PEREIRA, F. (2007). Biographies, Bollywood, Boom-boxes and Blenders : Domain Adaptation for Sentiment Classification. *In ACL 2007*, Prague, Czech Republic.
- CHOI, Y. et CARDIE, C. (2009). Adapting a Polarity Lexicon using Integer Linear Programming for Domain-Specific Sentiment Classification. *In EMNLP 2009*, pages 590–598, Singapore.
- DENECKE, K. (2009). Are SentiWordNet scores suited for multi-domain sentiment classification ? *In 4th International Conference on Digital Information Management (ICDIM 2009)*, pages 1–6.
- DRURY, B., TORGIO, L. et ALMEIDA, J. J. (2011). Guided self training for sentiment classification. *In Workshop on Robust Unsupervised and Semisupervised Methods in Natural Language Processing*.
- ESULL, A. et SEBASTIANI, F. (2006). SentiWordNet : A Publicly Available Lexical Resource for Opinion Mining. *In 5th Conference on Language Resources and Evaluation (LREC 2006)*.
- GINDL, S., WEICHELSELBRAUN, A. et SCHARL, A. (2010). Cross-Domain Contextualization of Sentiment Lexicons. *In 19th European Conference on Artificial Intelligence*, pages 771–776.
- HARB, A., DRAY, G., PLANTÉ, M., PONCELET, P., ROCHE, M. et TROUSSET, F. (2008). Détection d'opinion : Apprenons les bons adjectifs ! *In INFORSID'08 - Atelier FODOP'08*, pages 59–66.
- JIJKOUN, V., de RIJKE, M. et WEERKAMP, W. (2010). Generating focused topic-specific sentiment lexicons. *In 48th Annual Meeting of the Association for Computational Linguistics*, pages 585–594.
- LI, S. et ZONG, C. (2008). Multi-domain sentiment classification. *In 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies*, pages 257–260.
- SHAPIRE, R. E. et SINGER, Y. (2000). BoosTexter : A boosting-based system for text categorization. *Machine Learning*, 39(1):135–168.
- TABOADA, M., BROOKE, J., TOFILOSKI, M., VOLL, K. et STEDE, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37(2):267–307.
- TORRES-MORENO, J.-M., EL-BÈZE, M., BÉCHET, F. et CAMELIN, N. (2007). Comment faire pour que l'opinion forgée à la sortie des urnes soit la bonne ? Application au défi DEFT 2007. *In Atelier DEFT'07 - Plate-forme AFIA 2007*, Grenoble, France.
- WIEBE, J. et RILOFF, E. (2005). Creating subjective and objective sentence classifiers from unannotated texts. *In CICLing-2005*.