

Méthode faiblement supervisée pour l'extraction d'opinion ciblée dans un domaine spécifique

Romaric Besançon

CEA, LIST, Laboratoire Vision et Ingénierie des Contenus
CEA Saclay Nano-INNOV, PC no 173, 91191 Gif-sur-Yvette CEDEX
romaric.besancon@cea.fr

Résumé. La détection d'opinion ciblée a pour but d'attribuer une opinion à une caractéristique particulière d'un produit donné. La plupart des méthodes existantes envisagent pour cela une approche non supervisée, sans hypothèse sur les caractéristiques visées. Or, les utilisateurs ont souvent une idée *a priori* des caractéristiques sur lesquelles ils veulent découvrir l'opinion des gens. Nous proposons dans cet article une méthode pour une extraction d'opinion ciblée qui exploite cette information minimale sur les caractéristiques d'intérêt. Ce modèle s'appuie sur une segmentation automatique des textes, un enrichissement des données disponibles par similarité sémantique et une annotation de l'opinion par classification supervisée. Nous montrons l'intérêt de l'approche sur un cas d'étude dans le domaine des jeux vidéo.

Abstract.

A Weakly Supervised Approach for Aspect-Based Opinion Mining in a Specific Domain.

The goal of aspect-based opinion mining is to associate an opinion with fine-grain aspects of a given product. Most approaches designed in this purpose use unsupervised techniques, whereas the information of the desired targeted aspects can often be given by the end-users. We propose in this paper a new approach for targeted opinion detection that uses this minimal information, enriched using several semantic similarity measures, along with topical segmentation and supervised classification. We prove the interest of the approach on an evaluation corpus in the specific domain of video games.

Mots-clés : Analyse d'opinion, classification supervisée, similarité sémantique.

Keywords: Opinion analysis, classification, semantic similarity.

1 Introduction

L'analyse automatique de l'opinion est une tâche importante pour beaucoup d'entreprises qui souhaitent connaître l'accueil public de leur marque ou de leur produit et a, de ce fait, donné lieu à de nombreuses recherches ces dernières années (Pang & Lee, 2007; Liu, 2012). Dans toute sa généralité, la problématique de la détection d'opinion se compose de plusieurs tâches, qu'il peut être utile ou non de mettre en œuvre selon les applications visées : la détection de la présence ou non d'une opinion ; la classification de la polarité de l'opinion (positif, négatif, neutre) ; la classification de l'intensité de l'opinion ; l'identification de l'objet de l'opinion (ce sur quoi porte l'opinion) ; l'identification de la source de l'opinion (qui exprime l'opinion). Toutes ces tâches peuvent se pratiquer à différents niveaux : au niveau global du texte, au niveau très précis d'une caractéristique particulière ou à des niveaux intermédiaires comme la phrase ou le paragraphe. À ces différents aspects s'ajoute le problème du domaine thématique des textes étudiés. En effet, la détection de l'opinion s'appuie sur le sens et la polarité des mots, qui peuvent changer selon les domaines considérés (Marchand *et al.*, 2014).

La problématique à laquelle nous nous intéressons dans cet article est celle de la classification de l'opinion associée à certaines caractéristiques précises de l'objet d'étude, dans un domaine spécifique. Par exemple, dans le cas des hôtels, les différentes caractéristiques seraient le prix, la taille des chambres, la propreté etc. Dans le domaine de notre cas d'étude sur les jeux vidéo, ces caractéristiques sont le graphisme, le *gameplay*, la profondeur de jeu etc.

Dans les recherches précédentes sur le sujet, ce problème est le plus souvent abordé avec des approches non supervisées visant à extraire des couples d'objets et de caractéristiques de ces objets, puis à chercher à associer une polarité d'opinion à ces caractéristiques en utilisant des mesures d'association (Popescu & Etzioni, 2005). Ces associations entre objets et caractéristiques et entre caractéristiques et opinion s'appuient souvent sur des analyses complexes prenant en particu-

lier en compte les dépendances syntaxiques au niveau des phrases (Qiu *et al.*, 2011; Cataldi *et al.*, 2013). Par ailleurs, ces méthodes s'appuyant sur des analyses locales, des solutions doivent également être mises en œuvre pour pallier les problèmes de résolution de coréférences pour les cibles de l'opinion (Ding & Liu, 2010). Une autre famille d'approches pour répondre à ce problème s'appuie sur des modèles probabilistes génératifs de type LDA (*Latent Dirichlet Allocation*), pour essayer de modéliser ces différentes dépendances, entre l'objet et ses caractéristiques d'une part et avec entre ces caractéristiques et les expressions d'opinion d'autre part (Moghaddam & Ester, 2013; Xueke *et al.*, 2013; Titov & McDonald, 2008). Les caractéristiques extraites de cette façon sont des parties de l'objet, des propriétés ou des concepts liés et peuvent être relativement nombreuses : par exemple, (Hu & Liu, 2004) compte en moyenne près de 70 caractéristiques pour différentes classes d'objets (appareils photos, téléphones mobiles, lecteurs MP3 etc.)

Le problème que nous abordons ici est un peu différent : nous cherchons à nous appuyer sur une connaissance *a priori* des caractéristiques qui nous intéressent et sur lesquelles on veut connaître les opinions des utilisateurs. En effet, dans la pratique, les entreprises, qui sont spécialistes de leur domaine et de leurs produits, ont une assez bonne idée de ce sur quoi elles veulent avoir une opinion. D'une part, cette connaissance est une contrainte parce qu'elle cible les caractéristiques à chercher, à l'opposé des approches non supervisées ; d'autre part, même si cette connaissance est très limitée (elle peut se limiter, en fait, aux seuls noms de caractéristiques visées), il est possible de l'utiliser pour guider directement la détection ciblée de l'opinion.

Pour répondre à ce problème, une approche possible est d'appliquer les techniques complètement non supervisées citées plus haut pour extraire des caractéristiques sur lesquelles des opinions sont exprimées, puis d'essayer de relier ces caractéristiques extraites à celles visées. Ce type de démarche peut être qualifiée d'*ascendante* : on part d'informations locales dans le texte pour les relier aux caractéristiques visées. C'est, par exemple, l'approche envisagée dans la tâche *Aspect-based Sentiment Analysis* de SemEval 2014 (Pontiki *et al.*, 2014). Nous proposons ici une approche différente, de type *descendante* : on utilise directement la connaissance des caractéristiques visées pour identifier dans les documents les segments de textes relatifs à ces différentes caractéristiques et leur attribuer une opinion, en utilisant des techniques de classification d'opinion au niveau du segment de texte. Ce type d'approche descendante nous semble en effet plus robuste, en ce qu'elle s'appuie sur des segments de texte plus importants, à la fois pour la qualification de l'opinion et pour la reconnaissance du lien avec les caractéristiques visées, ce qui permet d'exploiter plus d'indices contextuels et éventuellement d'être moins dépendant du domaine. Cette méthode est néanmoins destinée à l'analyse de textes plutôt longs et structurés, *e.g.* des articles critiques, et n'est pas adaptée pour des textes très courts (*e.g.* des *tweets*).

Nous présentons notre approche plus en détails dans la section suivante, et en montrons une évaluation sur un corpus de critiques en anglais, dans le domaine des jeux vidéo, dans la section 3.

2 Approche proposée pour la détection d'opinion ciblée

2.1 Vue générale de l'approche

La problématique à laquelle nous voulons répondre est de n'utiliser qu'une information minimale (les noms des différentes caractéristiques) pour développer un système de reconnaissance automatique d'une opinion ciblée sur ces caractéristiques. L'approche que nous proposons est la suivante :

- un découpage automatique des documents en différents segments de texte ;
- l'association de ces segments à une des caractéristiques visées ;
- une annotation automatique de l'opinion effectuée indépendamment sur ces différents segments ;
- une annotation finale de l'opinion sur chaque caractéristique combinant les annotations sur chaque segment.

L'idée de cette approche est de mettre en place un système de reconnaissance de l'opinion sur les différentes caractéristiques visées de la façon la moins supervisée possible, c'est-à-dire en ne considérant que les données du problème telles qu'on l'a posé ci-dessus, sans autre connaissance structurée ou données d'apprentissage supplémentaires. En effet, pour la première étape, on pourrait envisager de faire une segmentation manuelle d'un ensemble de documents selon ces différentes caractéristiques pour entraîner, de façon automatique, un segmenteur spécialisé. De même, l'association d'une caractéristique à chaque segment de texte peut être facilitée par une ontologie du domaine associant un vocabulaire spécifique à chacune des caractéristiques. Or, même si on trouve de nombreuses ressources lexico-ontologiques pour les langues bien dotées, elles sont souvent généralistes et leur spécialisation sur un domaine (par extraction et enrichissement d'une partie des informations) demande souvent des interventions manuelles. Or, ces opérations manuelles de construction ou d'adaptation de ressources sont coûteuses et doivent être refaites pour chaque nouveau domaine ou chaque nouvelle caractéristique considérée. Dans l'optique d'avoir un système adaptable facilement à de nouveaux domaines (*i.e.* de façon

complètement automatique), nous choisissons de mettre en place, pour cette étude, une approche moins supervisée.

2.2 Reconnaissance automatique d'une opinion au niveau du texte

Dans l'approche que nous proposons, la détection de l'opinion se fait au niveau d'un segment de texte. Nous nous plaçons ici dans un paradigme de reconnaissance automatique de l'opinion par classification automatique, sans utilisation de connaissances extérieures, comme des lexiques d'opinion, toujours dans la logique que des ressources lexicales spécialisées ne sont pas forcément disponibles.

2.2.1 Prétraitement

Concernant le prétraitement des documents, nous avons volontairement choisi de ne garder qu'un prétraitement minimal. En effet, pour la reconnaissance automatique de l'opinion, il est important de garder les informations de flexion : l'opinion suggérée par « I love » (présent) n'est pas forcément la même que celle suggérée par « I loved » (au passé : si l'objet a été aimé, ce n'est plus forcément le cas). On se limite donc ici à faire une segmentation du texte en mots, en gardant tous les mots dans leur forme de surface originale (en opérant néanmoins une normalisation par la mise en minuscules des mots). Nous avons également utilisé un filtre par anti-dictionnaire (*stoplist*) pour supprimer les mots grammaticaux non porteurs d'opinion. Nous avons pour cela repris les *stopwords* fournis par (Blitzer *et al.*, 2007) avec le corpus *Multi-Domain Sentiment Dataset*.

Nous n'avons pas effectué de traitement linguistique spécifique pour prendre en compte la négation mais nous utilisons également des n-grams de mots pour la représentation des documents (de sorte que des expressions composées comme « *not bad* » pourront être prises en compte). Dans la pratique, nous considérerons des n-grams de taille 3.

2.2.2 Classification automatique de l'opinion

Nous optons dans cette étude pour une approche de détection de l'opinion par classification supervisée. Plus précisément, nous avons fait des tests en utilisant des classifieurs standard de type SVM (*Support Vector Machine*) (Vapnik, 1995) et AdaBoost (*Adaptive Boosting*).

Les SVM ont souvent été utilisés avec succès pour la tâche de la classification d'opinion (Pang *et al.*, 2002; Rushdi-Saleh *et al.*, 2011), parfois en s'appuyant sur des fonctions noyaux complexes ((Wu *et al.*, 2009) utilisent par exemple des *tree kernels*). Dans cette expérimentation, nous utilisons pour notre part des noyaux linéaires standards.

Le *boosting* (Schapire, 1999) est une technique d'apprentissage s'appuyant sur l'idée qu'on peut construire un classifieur efficace par une combinaison pertinente de classifieurs faibles. Dans le cas de l'analyse de texte, ces classifieurs faibles sont les mots ou des n-grams de mots. Cette méthode de classification statistique a déjà montré de bons résultats sur la tâche de détection d'opinion, par exemple lors de la campagne d'évaluation DEFT'07 (Torres-Moreno *et al.*, 2007), et présente par ailleurs l'avantage d'avoir un modèle relativement explicite : on identifie les mots ou n-grams les plus discriminants pour la classification, ce qui peut permettre d'acquérir automatiquement des ressources lexicales d'opinion dans le domaine considéré¹.

Les implémentations utilisées de ces classifieurs statistiques sont SVMlight (Joachims, 2002) et BoosTexter (Schapire & Singer, 2000).

2.3 Reconnaissance automatique d'une opinion ciblée

2.3.1 Segmentation automatique des documents

Pour la segmentation automatique des documents, nous utilisons la méthode *LCseg* de segmentation thématique automatique à base de chaînes lexicales (*Lexical Chain Segmenter* (Galley *et al.*, 2003)). Cette méthode de segmentation s'appuie sur le repérage de chaînes lexicales (suites de termes répétés, éloignés d'une distance inférieure à une valeur

1. Dans le cas d'un SVM linéaire, on peut également retrouver les traits les plus influents en utilisant une régularisation L1, mais l'approche est moins immédiate.

donnée) et l'utilisation d'une mesure de cohésion lexicale, définie sur des frontières données (changements de phrase) et dépendant des chaînes lexicales qui traversent ces frontières. La segmentation consiste à trouver les changements de phrases qui minimisent cette mesure de cohésion lexicale. Une des forces de cette méthode est son indépendance par rapport au domaine, ce qui en fait une bonne base générique de segmentation. De plus, cette technique de segmentation a montré de meilleures performances que des techniques de découvertes automatique de thèmes par LDA (*Latent Dirichlet Allocation*), par exemple pour la segmentation thématique d'emails (Joty *et al.*, 2010).

Pour évaluer l'intérêt d'une segmentation thématique des documents, nous comparons les résultats avec une segmentation effectuée sur la base des phrases (chaque phrase est considérée comme un segment).

2.3.2 Association d'une caractéristique à chacun des segments

Après avoir segmenté le texte selon des segments thématiquement cohérents, on cherche à associer une caractéristique à chaque segment. La seule information dont nous disposons est le nom des différentes caractéristiques. Comme cette seule information n'est certainement pas suffisante, la première étape consiste donc à l'enrichir en considérant d'autres mots associés.

La détermination de mots associés s'appuie sur une forme de similarité sémantique entre les mots. Dans cette étude, nous utilisons, pour établir cette similarité sémantique, des informations distributionnelles construites automatiquement. Plus précisément, nous avons testé plusieurs méthodes d'expansion :

- par les mots les plus fréquemment présents avec les mots désignant les caractéristiques (mots *co-occurents*) ;
- par des *voisins distributionnels* des caractéristiques, calculés selon un modèle de thésaurus distributionnel ;
- par des *voisins « représentationnels »* des caractéristiques, selon un modèle de représentations lexicales distribuées.

Pour la construction d'un thésaurus distributionnel, nous avons suivi la méthode de (Ferret, 2010) : dans un corpus donné, chaque mot est représenté par ses co-occurrences avec les autres mots ; une mesure de similarité entre les mots est alors définie par la similarité entre leurs profils de co-occurrence (mesurée par le cosinus entre les vecteurs représentant ces profils). Cette méthode est paramétrée par la taille de la fenêtre des mots co-occurents, *i.e.* nombre de mots pleins (noms, verbes ou adjectifs) considérés comme co-occurents de chaque côté du mot considéré (une fenêtre de taille 1 correspond donc à considérer dans le profil de co-occurrence un mot à gauche et un mot à droite). La taille de cette fenêtre change en général la nature de la similarité sémantique : une fenêtre de petite taille a tendance à favoriser les similarités de type *synonyme* (les mots partagent des contextes locaux proches, e.g. ils apparaissent comme objets des mêmes verbes), alors qu'une fenêtre de plus grande taille génère des regroupements plus thématiques.

Les représentations lexicales distribuées sont construites selon la méthode de (Mikolov *et al.*, 2013), en utilisant son outil disponible *word2vec*². L'idée de cette méthode est d'utiliser un réseau de neurones pour apprendre une représentation vectorielle des mots qui arrive à capter une similarité sémantique entre les mots, fondée sur leur contexte. Deux approches sont en fait proposées : l'une, dite en sac-de-mots continu (*continuous bag-of-words* ou CBOW), cherche à maximiser la probabilité d'un mot en fonction de son contexte, alors que l'autre (*skip-gram*) cherche, au contraire, à prédire le contexte sachant le mot. En plus du modèle utilisé, cette approche utilise également d'autres paramètres, dont la taille de la fenêtre de contexte considérée et la taille du vecteur de représentation des mots.

La qualité des données ainsi construites (co-occurrences simples ou thésaurus distributionnels) dépend de la taille de la collection : dans notre cas, nous avons utilisé les collections spécialisées sur les jeux vidéo dont nous disposons, pour un corpus total d'environ 500 000 mots³. Pour les représentations lexicales distribuées, on comparera les résultats à ceux obtenus avec des voisins utilisant des représentations pré-calculées, fournies avec l'outil, sur un corpus généraliste très important (corpus *Google News*, de 3 milliards de mots).

Pour ces différentes méthodes, un score $w_{\text{assoc}}(c, t)$ est attribué à l'association du nom de la caractéristique c considérée avec chaque terme t utilisé pour son expansion : dans le cas des co-occurents, ce score est lié à la fréquence de co-occurrence, dans le cas des voisins sémantiques, ce score est lié à la distance entre les profils de co-occurrence ou les représentations lexicales distribuées.

Un poids d'association $w(s, c)$ d'une caractéristique c à un segment $s = \{t_1, \dots, t_n\}$ est alors défini par la somme, pour

2. <https://code.google.com/p/word2vec/>

3. Nous utilisons donc, pour partie, notre corpus de test dans les données sur lesquelles la construction des réseaux de co-occurrences est faite, ce qui présente en fait un biais dans l'évaluation. Néanmoins, même en intégrant ces données, le corpus reste relativement petit pour ce type de tâche, et nous avons donc décidé de les conserver. Idéalement, d'autres collections du domaine, non annotées, devraient être collectées pour construire ces ressources sémantiques.

chacun des termes du segment, du poids d'association de ce terme avec la caractéristique, selon le modèle d'expansion choisi. On attribue alors à chaque segment s la caractéristique $\text{car}(s)$ de plus haut poids :

$$\text{car}(s) = \underset{c}{\operatorname{argmax}} \sum_{t_i \in s} w_{\text{assoc}}(c, t_i)$$

2.3.3 Annotation automatique de l'opinion associée aux caractéristiques

Pour déterminer l'opinion de chacun des segments, nous utilisons simplement un modèle de reconnaissance de l'opinion générale, présentée en section 2.2, entraîné sur des textes entiers annotés en opinion et appliqué sur le segment de texte. Nous obtenons ainsi, pour un segment s , un score pour chaque décision d'opinion (positive/négative), notés $w_{op}(\text{pos}, s)$ et $w_{op}(\text{neg}, s)$.

Si on note $S(c)$ l'ensemble des segments liés à la caractéristique c (i.e. $S(c) = \{s | \text{car}(s) = c\}$), l'annotation globale de l'opinion sur c est alors obtenue en faisant simplement la moyenne des scores d'opinion sur chacun des segments de cet ensemble et en prenant la décision de l'opinion (positive/négative) sur la base de ces moyennes.

$$\text{opinion}(c) = \begin{cases} \text{positive si } w_{op}(\text{pos}, c) > w_{op}(\text{neg}, c) \\ \text{négative si } w_{op}(\text{neg}, c) > w_{op}(\text{pos}, c) \end{cases}$$

$$\text{avec } w_{op}(x, c) = \frac{1}{|S(c)|} \sum_{s \in S(c)} w_{op}(x, s)$$

3 Évaluation

Les expériences pour l'évaluation de la méthode proposée ont été réalisées sur plusieurs corpus de documents en anglais⁴, dans le domaine des jeux vidéo. Nous avons retenu les caractéristiques suivantes : le graphisme (*Graphics*), le son (*Sound*), le *Gameplay*, la profondeur du jeu (*Depth*), la présentation (*Presentation*). Ces caractéristiques nous ont été suggérées par les systèmes de notations adoptés par les critiques de différents sites de jeux vidéo, en particulier le site *videogamesdaily.com*, à partir duquel nous avons construit nos données de référence, mais d'autres sites utilisent des critères similaires : par exemple, sur les archives du site de critiques en français *jeuxvideo.com*, on trouve une séparation de la note selon des caractéristiques proches : *Graphismes*, *Jouabilité*, *Durée de vie*, *Bande son*.

3.1 Corpus d'évaluation

3.1.1 Corpus MDS

Le corpus *Multi-Domain Sentiment Dataset* a été utilisé pour la première fois par (Blitzer *et al.*, 2007). C'est une collection de critiques en anglais, pour différents produits dans des domaines différents, récupérées sur le site Amazon. Ce corpus a été utilisé en particulier pour étudier la détection automatique de l'opinion sur des textes en fonction du domaine. Ces critiques sont des petits textes (en moyenne une centaine de mots) associés à une indication de la satisfaction de l'utilisateur sous la forme d'une note sur 5 (nombre d'étoiles). Pour bien distinguer les critiques positives et négatives, les documents ayant une note de 4 ou 5 sont considérés positifs, ceux ayant une note de 1 ou 2 négatifs (les documents ayant une note de 3 sont ignorés). Le nombre de documents visé pour chaque domaine est de 1000 documents positifs et autant de négatifs, mais certains domaines sont moins représentés et la collection contient au final 38548 documents.

Dans le cadre de cette étude, nous nous sommes focalisés sur le domaine des jeux vidéo. Nous avons donc extrait du corpus *Multi-Domain Sentiment Dataset* un sous-ensemble de critiques portant sur les jeux vidéo. Plus précisément, ces critiques ont été prises dans le domaine « *Computer & Video Games* » et ont été filtrées sur la base de mots clés pour supprimer les produits liés à l'informatique qui ne sont pas des jeux vidéo. On a ainsi retenu 1229 critiques, dont 848 positives et 381 négatives. Ce corpus est noté MDS-JV dans les résultats des expériences.

4. Les techniques mises en œuvre ne s'appuyant pas sur des ressources explicites, elles sont indépendantes de la langue : il est néanmoins nécessaire, pour traiter une langue différente, de disposer au moins d'un corpus annoté en opinion dans cette langue, soit du domaine considéré, soit d'un domaine général.

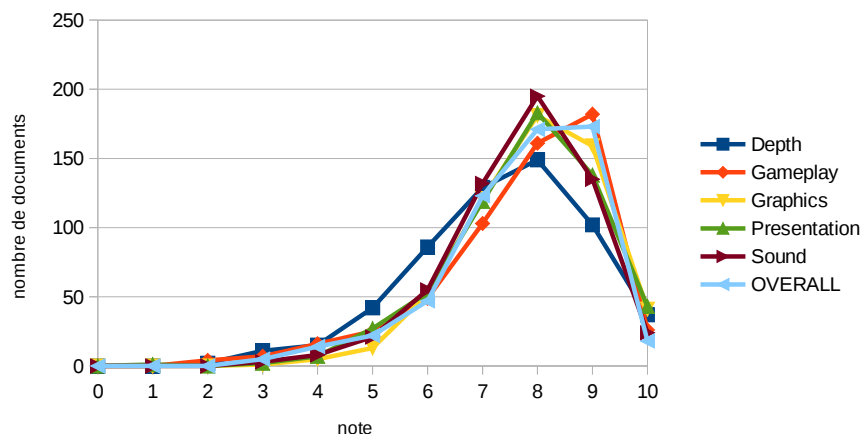


FIGURE 1 – Distribution des notes dans le corpus *videogamesdaily* selon les différentes caractéristiques (les notes sont arrondies à la valeur entière la plus proche).

3.1.2 Corpus *videogamesdaily*

Pour l'évaluation de la détection d'opinion ciblée, nous avons construit un corpus à partir des critiques des archives du site *videogamesdaily.com*, dans lequel les articles de critiques de jeux vidéo se concluent par une grille d'appréciation selon différents points : le graphisme, le son, le *gameplay*, la profondeur, la présentation, avec, pour chacun de ces points, une note décimale comprise entre 1 et 10.

Nous avons collecté et reformaté ces archives pour former une collection annotée pour chacune des caractéristiques. Nous avons ainsi obtenu 573 articles. Les documents de ce corpus étant des articles critiques d'un site spécialisé, ils sont plus détaillés que des critiques Amazon et ont en particulier une taille bien plus importante que les documents du corpus MDSD : 460 mots en moyenne contre 200 mots pour le corpus MDSD-JV.

En ce qui concerne la distribution des notes, présentée à la figure 1, selon les différentes caractéristiques considérées, on constate une sur-représentation des notes positives (qui est également présente dans le corpus initial des critiques Amazon) et également un décalage de l'échelle de notation : les notes de 1 à 3 ne sont quasiment pas utilisées, les courbes sont plutôt centrées sur des notes moyennes autour de 8. De façon similaire au corpus extrait d'Amazon, on choisit comme exemples de documents négatifs des documents ayant des notes de 6 ou moins et des documents ayant des notes de 9 à 10 comme exemples de documents positifs. Les notes entre 6.1 et 8.9 sont ignorées. Le nombre de documents positifs et négatifs ainsi obtenus, pour chacune des caractéristiques, est présenté dans le tableau 1. Notons que, si l'on a entre 199 et 257 documents pour chacune des caractéristiques, ces documents couvrent au total 462 documents différents du corpus (ce qui signifie qu'il n'y a pas énormément de recoupement entre les différents ensembles : en pratique, seuls 49 documents sont communs à tous les ensembles). Par ailleurs, dans la plupart des cas, les critiques ont souvent des notes sur les différentes caractéristiques qui vont dans le même sens, mais on compte néanmoins environ 12% des documents qui ont au moins une caractéristique négative et une caractéristique positive.

	nb docs positifs	nb docs négatifs
<i>Gameplay</i>	156	101
<i>Depth</i>	100	155
<i>Presentation</i>	135	90
<i>Graphics</i>	142	70
<i>Sound</i>	113	86

TABLE 1 – Nombre de documents positifs et négatifs selon chacune des caractéristiques considérées.

work minutes what#was rent repetitive
 bought#game cannot didn't boring nothing
 believe waste was don't 360 not#much#fun
 money gave poor completely not horrible
 unless look needs why returned worst i#got
 terrible

note music also never liked light d rocks variety
 different knows live 7 amazing must own
 if#you#don't love inside shooter ps2 enough
 close side who best fun better#than well
 original

FIGURE 2 – Les mots discriminants pour l’opinion, appris par BoosTexter sur le corpus MDSD-JV (les mots en rouge sont négatifs, les mots en vert positifs, la taille des mots est proportionnelle à l’importance accordée à ces mots par le modèle)

3.2 Résultats de classification d’opinion générale

Les premières expérimentations évaluent la qualité de la détection de l’opinion au niveau du texte sur ces collections. Pour chaque collection, les tests ont été réalisés par validation croisée (*10-fold*). Les résultats présentés dans le tableau 2 sont des moyennes obtenues sur les 10 essais. Les mesures d’évaluation utilisées pour la détection de l’opinion sont l’exactitude des résultats (*accuracy*), définie par le rapport des annotations correctes sur le nombre total d’annotations, ainsi que les précision/rappel/f-score, calculées comme des moyennes des scores pour les opinions positives et négatives.

	BoosTexter		SVM	
	videogamesdaily	MDSD-JV	videogamesdaily	MDSD-JV
Exactitude	77,98%	90,16%	77,55%	88,72%
F-score	77,49%	90,03%	77,12%	88,59%
Précision	78,96%	90,56%	78,49%	89,42%
Rappel	78,62%	90,17%	79,43%	88,71%

TABLE 2 – Évaluation de la reconnaissance automatique de l’opinion par apprentissage automatique

On voit dans ces résultats que les deux types de classifieurs donnent des résultats relativement bons et comparables, même si les résultats obtenus par BoosTexter sont un peu meilleurs que ceux obtenus avec les SVM. Les résultats sont meilleurs pour le corpus MDSD-Jeux vidéo, ce qui n’est pas étonnant étant donné que c’est le corpus de taille la plus importante (le corpus *videogamesdaily* est relativement limité pour des expérimentations par apprentissage automatique).

Un avantage de l’approche par Boosting est de construire un modèle qui s’appuie explicitement sur les mots comme classifieurs faibles et permet donc de visualiser le vocabulaire utilisé par le classifieur pour différencier les textes positifs et négatifs. La figure 2 présente par exemple les 30 mots les plus discriminants pour le corpus MDSD-Jeux vidéo. On voit dans ces listes des mots qui sont porteurs d’opinion (*horrible*, *terrible*, *repetitive*, *not#much#fun* pour le lexique négatif, *amazing*, *best*, *fun* pour le lexique positif), ce qui montre qu’on arrive à capter de façon automatique des expressions d’opinion. On trouve également d’autres mots spécifiques au domaine (*e.g.* *360* a une connotation négative alors que *PS2* a une connotation positive dans les critiques considérées).

Afin de tester la capacité des classifieurs à construire un modèle suffisamment général, nous entraînons des modèles sur un corpus et les testons sur l’autre. Ces expérimentations ont seulement été faites avec le modèle BoosTexter, qui donnait plutôt de meilleurs résultats sur les corpus pris isolément. Par ailleurs, (Mansour *et al.*, 2013) a montré qu’un classifieur entraîné sur un grand nombre de données de différents domaines permettait parfois d’avoir de meilleurs résultats qu’un classifieur entraîné sur le domaine visé. Dans cette optique, nous reportons également les résultats en entraînant le classifieur sur la totalité des données du corpus MDSD et en faisant le test sur nos deux corpus spécialisés. Les résultats de ces tests croisés, en exactitude, sont présentés dans le tableau 3 (les résultats des corpus sur eux-mêmes sont ceux obtenus en validation croisée, alors que pour les résultats croisés, la totalité des documents du corpus source a été utilisée pour l’apprentissage et la totalité du corpus cible pour le test).

On remarque à nouveau dans ces résultats, l’impact de la taille du corpus, le modèle appris sur le corpus *videogamesdaily* menant à un modèle plus spécialisé qui ne s’adapte pas bien sur d’autres collections. Par contre, pour le corpus MDSD-JV, le modèle appris permet d’obtenir des résultats intéressants sur un autre corpus de nature différente.

Le dernier résultat montre également qu’on arrive à avoir une meilleure généralisation en apprenant sur une grande quan-

		testé sur	
		videogamesdaily	MDS-DJV
appris sur	videogamesdaily	77,98%	52,62%
	MDS-DJV	66,67%	90,16%
	MDS-D	77,98%	83,73%

TABLE 3 – Résultats d’exactitude en entraînant sur un corpus et en testant sur un autre corpus.

waste not#buy return bad **disappointing**
 money broke useless **boring** not#worth **great** solid wonderful easy favorite my#only love
worst junk way#too horrible ok returned highly job pleased not#too delicious perfectly
 beware nothing poorly **poor** disappointed perfect excellent better#than great#product
 not#good not#recommend unfortunately amazing awesome i#like price best loves fun
disappointment awful don't#buy **not** your#money grill not#disappointed nice comfortable you#not enjoy
terrible

FIGURE 3 – Les mots discriminants pour l’opinion, appris par BoosTexter sur le corpus MDS-D complet (les mots en rouge sont négatifs, les mots en vert positifs, la taille des mots est proportionnelle à l’importance accordée à ces mots par le modèle).

tité de données, même en incluant des documents de domaines différents. Il faut noter que d’autres études, comme (Garcia-Fernandez *et al.*, 2014), montrent, à l’opposé, qu’à taille de corpus égal, un corpus homogène reste préférable. (Mansour *et al.*, 2013) rapporte des résultats moyens de l’ordre de 90 % sur chacun des domaines du corpus MDS-D pris séparément. On a donc ici des résultats un peu moins bons, mais qui restent relativement comparables. Si l’on regarde les mots discriminants appris sur ce corpus (présentés à la figure 3), on constate qu’on apprend en effet des mots porteurs d’opinion plus génériques, sans mots spécifiques du domaine.

3.3 Résultats de classification d’opinion ciblée

Nous évaluons la méthode proposée pour la classification d’opinion ciblée sur le corpus *videogamesdaily*, qui contient des annotations de référence selon les différentes caractéristiques. Comme les annotations de référence ne sont conservées que pour ces scores inférieurs à 6 ou supérieurs à 9, les documents n’ont pas d’annotations pour toutes les caractéristiques. Réciproquement, il est possible que la procédure d’annotation ne produise pas d’annotation sur un document pour une caractéristique donnée. Les scores d’exactitude présentés sont, de ce fait, calculés en considérant seulement l’intersection des annotations de référence avec les annotations automatiques. De ce fait, on rapporte également, pour les différents résultats, le nombre d’annotations qui sont prises en compte pour l’évaluation.

La segmentation thématique des documents par LCseg produit entre 1 et 22 segments par document avec une moyenne d’environ 9 segments. Les segments font en moyenne 86 mots (après suppressions des mots outils). Une segmentation par phrase mène à un découpage moyen des documents en 54 segments, d’une taille moyenne de 15 mots.

Pour les résultats, nous avons plusieurs paramètres à étudier :

- l’influence de la segmentation, en phrases ou en segments thématiques ;
- l’influence du modèle d’association des mots pour l’expansion : mots co-occurents, voisins sémantiques par thésaurus distributionnels ou par représentation lexicale distribuée ;
- la taille de la fenêtre pour la prise en compte du contexte ;

Le tableau 4 présente les résultats obtenus, en termes d’exactitude, sur les différentes caractéristiques considérées, pour un découpage en phrases ou en segments thématiques, en utilisant les voisins sémantiques calculés sur la base d’un thésaurus distributionnel construit en utilisant plusieurs taille de fenêtre de voisinage (en prenant 1,2,5 ou 10 mots de contextes).

Ces résultats mettent en évidence l’intérêt de l’utilisation d’une segmentation thématique par rapport à une simple segmentation en phrases pour obtenir une plus grande précision de détection d’opinion. L’amélioration observée peut également être partiellement due au fait que les segments thématiques sont de taille plus importante et supportent donc plus d’élé-

	segments thématiques				phrases			
	w1	w2	w5	w10	w1	w2	w5	w10
Depth	55,7%	65,2%	58,4%	62,2%	45,7%	52,0%	49,6%	46,8%
Gameplay	69,9%	70,6%	68,8%	71,0%	61,7%	62,1%	60,9%	62,1%
Graphics	68,4%	67,9%	74,7%	69,7%	70,1%	67,3%	69,2%	67,8%
Presentation	66,3%	68,2%	63,3%	61,3%	68,3%	61,5%	63,9%	63,8%
Sound	69,2%	71,7%	71,6%	70,5%	60,8%	60,9%	66,7%	63,8%
average	65,9%	68,7%	67,4%	66,9%	61,3%	60,8%	62,1%	60,9%
annotations	706	680	697	694	1059	1052	1043	1055

TABLE 4 – Résultats (exactitude) pour la détection d’une opinion ciblée, en comparant un découpage en phrases ou en segments thématiques, pour des voisins distributionnels calculés avec différentes tailles de contexte.

	Co-occurents				Voisins
	w1	w2	w5	w10	w2
Depth	57,7%	47,9%	50,6%	59,2%	65,2%
Gameplay	56,4%	57,4%	62,7%	56,9%	70,6%
Graphics	53,3%	60,0%	64,9%	58,6%	67,9%
Presentation	65,5%	69,0%	67,7%	71,5%	68,2%
Sound	62,4%	66,3%	73,5%	70,1%	71,7%
average	59,0%	60,1%	63,9%	63,3%	68,7%
annotations	732	742	668	612	680

TABLE 5 – Résultats (exactitude) pour la détection d’une opinion ciblée : comparaison d’expansion sémantique entre co-occurents et voisins distributionnels.

ments pour l’application du modèle de détection de l’opinion. En effet, l’attribution d’une opinion à un texte fonctionne en général d’autant mieux que le texte contient plus de mots susceptibles de porter une information de polarité d’opinion : si l’on attribue à chaque document une opinion calculée comme la moyenne des opinions détectées pour chacun des segments qui le composent, on obtient un score d’exactitude sur l’annotation globale des documents qui est de 66,7 % pour le découpage en segments thématiques et de 59,8 % pour le découpage en phrases. En comparant avec le score obtenu à partir de la totalité des documents (77,98 %), on remarque effectivement une baisse de performance due, vraisemblablement, à l’application de la détection d’opinion sur des documents de plus petite taille.

D’autre part, cette différence entre segments et phrases s’explique aussi par le fait que plus d’annotations sont prises en compte pour l’évaluation avec la segmentation en phrases : en effet, dans ce cas, la probabilité qu’au moins une des phrases soit associée à une caractéristique est plus importante, et on a donc globalement moins de caractéristiques pour lesquelles aucune annotation n’est fournie.

Par ailleurs, on remarque un comportement globalement meilleur pour une taille de fenêtre de 2, ce qui semble un bon compromis entre une mesure plus proche de la notion de synonymie et une fenêtre suffisamment large pour prendre en compte la petite taille du corpus sur lequel les co-occurrences sont apprises.

Le tableau 5 présente les résultats en comparant les meilleurs résultats obtenus avec les voisins sémantiques (sur un contexte de deux mots) à ceux obtenus avec les mots co-occurents pour l’attribution des caractéristiques à chaque segment, pour différentes tailles de fenêtres de co-occurrence (les résultats sont obtenus avec la segmentation thématique). Ce tableau montre que l’utilisation de voisins sémantiques donne globalement de meilleurs résultats que l’utilisation des simples co-occurents. On constate néanmoins que pour certaines des caractéristiques, les résultats peuvent être meilleurs avec les co-occurents. En particulier, pour la caractéristique « *Presentation* », les résultats sont de façon générale moins bons avec les voisins sémantiques : cette tendance était également vérifiée dans le tableau précédent, avec le découpage en phrases, ce qui montre que, pour cette caractéristique, les voisins sémantiques ne fournissent pas une bonne expansion (ce qui peut s’expliquer par la nature particulièrement polysémique du mot « *presentation* ») . De façon générale, néanmoins, les valeurs moyennes sont plus élevées avec les voisins sémantiques.

De façon complémentaire, on peut voir des différences entre les attributions des caractéristiques aux différents segments selon qu’on utilise des co-occurents ou des voisins sémantiques. Le tableau 6 montre le nombre de segments attribués à chacune des catégories pour les deux modèles. On remarque plusieurs choses en observant ces distributions : d’une

	segments thématiques				phrases			
	voisins		co-occurents		voisins		co-occurents	
	w2	w10	w2	w10	w2	w10	w2	w10
Depth	183	243	1805	1244	1460	2130	7165	2183
Gameplay	3107	3110	171	310	12471	12878	1027	711
Graphics	1213	1242	190	69	6892	7249	983	232
Presentation	183	171	1356	985	1508	1446	5638	1758
Sound	405	378	1407	303	2273	2186	5826	610
No category	216	163	378	2396	6232	4947	10197	25342

TABLE 6 – Nombre de segments/phrases attribués à chaque caractéristique, selon le modèle d’expansion sémantique choisi.

part, l’utilisation de voisins sémantiques ou de co-occurents change l’attribution des caractéristiques aux segments, en favorisant les caractéristiques *Gameplay* et *Graphics* avec les voisins sémantiques alors qu’elles sont minoritaires en considérant les co-occurents ; d’autre part, l’augmentation de la taille de la fenêtre de co-occurrence tend à accentuer cette tendance. Enfin, ces deux observations sont conservées que ce soient avec les segments thématiques ou les phrases. Ces tendances peuvent être dues aux différences de nature des mots, *Gameplay* et *Graphics* étant fortement liés au domaine thématique alors que les autres mots ont aussi des sens plus généraux.

Le tableau 7 contient les résultats obtenus en utilisant les représentations lexicales distribuées pour le calcul des voisins sémantiques, d’une part en calculant ces représentations sur notre corpus spécialisé dans le domaine des jeux vidéo, pour différents voisinages considérés (des contextes de tailles 1,2,3,5 et 10 ont été testés)⁵ et, d’autre part, en utilisant les représentations lexicales pré-calculées sur le corpus général *Google News*, beaucoup plus important.

	corpus Jeux Vidéo								corpus Google	
	segments thématiques				phrases				segments	phrases
	w1	w2	w5	w10	w1	w2	w5	w10		
Depth	60,6%	61,6%	61,1%	54,4%	57,2%	53,0%	50,3%	49,8%	60,8%	53,8%
Gameplay	66,7%	67,7%	71,2%	70,0%	62,0%	61,7%	63,8%	65,8%	68,1%	68,8%
Graphics	77,8%	72,5%	70,0%	70,1%	71,9%	69,7%	67,3%	67,3%	82,3%	74,6%
Presentation	67,3%	66,1%	54,8%	58,0%	63,9%	63,9%	63,3%	67,7%	66,2%	63,6%
Sound	69,4%	73,8%	74,5%	72,9%	65,0%	66,3%	65,7%	63,8%	65,1%	65,4%
average	68,3%	68,4%	66,3%	65,1%	64,0%	62,9%	62,1%	62,9%	68,5%	65,2%
annotations	551	661	651	698	801	929	978	1037	632	776

TABLE 7 – Résultats (exactitude) pour la détection d’une opinion ciblée, avec utilisation des représentations lexicales distribuées.

Ce tableau montre qu’on peut obtenir, avec ce modèle, des résultats moyens comparables avec l’approche par thésaurus distributionnel et confirme l’intérêt de la segmentation thématique. On remarque néanmoins que les écarts entre les scores pour les différentes caractéristiques sont plus importants qu’avec les voisins distributionnels (écart-type de 4,95 comparé à un écart-type de 2,37 pour les voisins distributionnels). On remarque aussi qu’avec ce modèle, l’augmentation de la taille de la fenêtre tend à diminuer la qualité des résultats. Pour les résultats avec les représentations générales, on remarque à nouveau avec des écarts de scores plus importants et en particulier un score bien supérieur pour la caractéristique *Graphics*, qui a peut-être une distribution dans ce corpus particulièrement liée au domaine des jeux vidéos. Des tests supplémentaires devraient également être menés pour prendre ces représentations générales comme point de départ de l’entraînement sur le corpus spécialisé. Ces résultats montrent néanmoins, de façon intéressante, que pour un domaine spécifique, on n’a pas forcément besoin d’une collection de documents très importante pour construire des modèles intéressants. A contrario, cela montre aussi qu’avec un corpus très important pour entraîner un modèle, on peut construire un modèle qui aura des performances acceptables, même en domaine spécifique.

En ce qui concerne l’établissement des voisins sémantiques, le tableau 8 présente, pour illustration, les 10 voisins sémantiques

5. Ces résultats sont obtenus avec le modèle *CBOW* et une taille de représentation de 500, en utilisant le *negative sampling*. Nous avons fait des tests avec les différents paramètres (modèle, taille de la représentation), dont nous ne présentons pas tous les résultats faute de place. Les résultats présentés offrent un bon compromis entre le nombre d’annotations et la qualité des annotations.

tiques les plus proches pour chacune des caractéristiques considérées, obtenus d’une part avec un thésaurus distributionnel (fenêtre de taille 2) et d’autre part avec les représentations lexicales distribuées (modèle CBOW, fenêtre de taille 2 et taille de représentation de 500).

	voisins distributionnels	voisins word2vec
Depth	<i>flavour, layer, ton, challenge, amount, something, scope, strategy, variety, backdrop</i>	<i>breadth, wealth, layer, originality, longevity, astounding, periphery, fidelity, limitless, several</i>
Gameplay	<i>experience, play, control, mechanic, combat, storey, storyline, action, predecessor, meat</i>	<i>subplot, game, shoehorned, sorely, mechanics, multiplayer, monotony, superficial, unoriginal, fundamental</i>
Graphics	<i>background, animation, texture, detail, game, environment, model, version, great, engine</i>	<i>horrendous, replayability, outdated, psx, pros, portrait, summary, integrated, youre, looker</i>
Presentation	<i>visual, package, highlight, prowess, mixture, flair, standpoint, tt, quality, acting</i>	<i>accompaniment, aural, markedly, fidelity, prowess, inspiring, blemish, assured, ditty, nxc</i>
Sound	<i>effect, voice, music, soundtrack, guncon, noise, lighting, chaos, robot, sublime</i>	<i>ambient, particle, narration, music, stereo, crisp, 5.1, echo, sharp, echoes</i>

TABLE 8 – Voisins sémantiques des noms des caractéristiques considérées.

3.3.1 Limitations

Nous avons indiqué, dans la présentation du corpus, que les critiques contenant à la fois des jugements positifs et négatifs sur les différentes caractéristiques sont relativement peu nombreuses (environ 12% des documents). L’approche naïve consistant à donner à chacune des caractéristiques l’orientation d’opinion globale du document donne donc, statistiquement, de bons résultats. En pratique, on obtient avec cette approche *baseline* un score moyen de 73,54%, qui est en effet supérieur aux scores obtenus ici. Néanmoins, si l’on restreint l’évaluation aux seuls documents contenant effectivement les deux orientations, on obtient un score moyen de 57.47% avec cette approche naïve alors qu’on a un score de 67,47% pour la meilleure approche proposée dans cet article. Ce type d’approche est donc à privilégier si les opinions exprimées sont effectivement variées selon les caractéristiques dans les données considérées.

4 Conclusion

Nous présentons dans cet article une approche pour la détection d’opinion ciblée qui, à l’inverse des méthodes existantes non supervisées, utilise une faible supervision sur la nature des caractéristiques visées. Cette information minimale est étendue de façon automatique à l’aide de similarités sémantiques et est exploitée dans un cadre de classification automatique de l’opinion. L’approche est évaluée dans le domaine des jeux vidéo et donne des résultats intéressants, avec presque 70 % de bonne détection de l’opinion sur chacune des caractéristiques visées. Ces résultats sont d’autant plus prometteurs qu’ils s’appuient sur une méthode très peu supervisée, demandant très peu de connaissances externes. Concernant la similarité sémantique utilisée, on montre que des approches à base de thésaurus distributionnels ou de représentations lexicales distribuées donnent des résultats moyens comparables, même si les approches par représentations distribuées semblent donner des résultats moins réguliers sur les différentes caractéristiques.

Parmi les développements futurs de cette étude, on peut envisager d’enrichir les traits pris en compte pour la détection de l’opinion, en intégrant par exemple des lexiques d’opinion, même généralistes, dans les modèles. Il faudrait également comparer l’approche proposée aux approches non supervisées, par exemple utilisées dans la campagne SemEval, sur les mêmes données. Enfin, une autre extension possible de ce travail serait de déterminer si le fait de disposer d’une évaluation de l’opinion pour chacune des caractéristiques permet d’améliorer la détermination de l’opinion globale.

Références

- BLITZER J., DREDZE M. & PEREIRA F. (2007). Biographies, bollywood, boom-boxes and blenders : Domain adaptation for sentiment classification. In *Association for Computational Linguistics*, Prague, Czech Republic.
- CATALDI M., BALLATORE A., TIDDI I. & AUFAURE M.-A. (2013). Good location, terrible food : detecting feature sentiment in user-generated reviews. *Social Network Analysis and Mining*, 3(4), 1149–1163.

- DING X. & LIU B. (2010). Resolving object and attribute coreference in opinion mining. In *Proceedings of COLING '10*, p. 268–276, Stroudsburg, PA, USA : Association for Computational Linguistics.
- FERRET O. (2010). Testing semantic similarity measures for extracting synonyms from a corpus. In *Proceedings of LREC'10*.
- GALLEY M., MCKEOWN K., FOSLER-LUSSIER E. & JING H. (2003). Discourse segmentation of multi-party conversation. In *Proceedings of ACL '03*, p. 562–569, Stroudsburg, PA, USA : Association for Computational Linguistics.
- GARCIA-FERNANDEZ A., FERRET O. & DINARELLI M. (2014). Evaluation of different strategies for domain adaptation in opinion mining.
- HU M. & LIU B. (2004). Mining opinion features in customer reviews. In *Proceedings of the 19th National Conference on Artificial Intelligence, AAAI'04*, p. 755–760 : AAAI Press.
- JOACHIMS T. (2002). *Learning to Classify Text Using Support Vector Machines – Methods, Theory, and Algorithms*. Kluwer/Springer.
- JOTY S., CARENINI G., MURRAY G. & NG R. T. (2010). Exploiting conversation structure in unsupervised topic segmentation for emails. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, EMNLP '10*, p. 388–398, Stroudsburg, PA, USA : Association for Computational Linguistics.
- LIU B. (2012). *Sentiment Analysis and Opinion Mining*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers.
- MANSOUR R., REFAEI N., GAMON M., ABDUL-HAMID A. & SAMI K. (2013). Revisiting the old kitchen sink : Do we need sentiment domain adaptation ? In *Proceedings of RANLP 2013*, p. 420–427.
- MARCHAND M., BESANÇON R., MESNARD O. & VILNAT A. (2014). Influence des marqueurs multi-polaires dépendant du domaine pour la fouille d'opinion au niveau du texte. In *Proceedings of TALN 2014*, p. 1–12.
- MIKOLOV T., CHEN K., CORRADO G. & DEAN J. (2013). Efficient estimation of word representations in vector space. In *Proceedings of Workshop at ICLR*.
- MOGHADDAM S. & ESTER M. (2013). The flda model for aspect-based opinion mining : Addressing the cold start problem. In *Proceedings of the 22Nd International Conference on World Wide Web, WWW '13*, p. 909–918.
- PANG B. & LEE L. (2007). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, **2**.
- PANG B., LEE L. & VAITHYANATHAN S. (2002). Thumbs up ? : Sentiment classification using machine learning techniques. In *Proceedings of EMNLP '02*, p. 79–86, Stroudsburg, PA, USA : Association for Computational Linguistics.
- PONTIKI M., GALANIS D., PAVLOPOULOS I., PAPAGEORGIOU H., ANDROUTSOPOULOS I. & MANANDHAR S. (2014). SemEval 2014 Task 4 : Aspect Based Sentiment Analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014) at (COLING 2014)*, p. 27–35, Dublin, Ireland.
- POPESCU A.-M. & ETZIONI O. (2005). Extracting product features and opinions from reviews. In *Proceedings of HLT '05*, p. 339–346, Stroudsburg, PA, USA : Association for Computational Linguistics.
- QIU G., LIU B., BU J. & CHEN C. (2011). Opinion word expansion and target extraction through double propagation. *Computational Linguistics*, **37**, 9–27.
- RUSHDI-SALEH M., MARTÍN-VALDIVIA M. T., RÁEZ A. M. & LÓPEZ L. A. U. (2011). Experiments with SVM to classify opinions in different domains. *Expert Syst. Appl.*, **38**(12), 14799–14804.
- SCHAPIRE R. E. (1999). A brief introduction to boosting. In *Proceedings of IJCAI'99*, p. 1401–1406, San Francisco, CA, USA : Morgan Kaufmann Publishers Inc.
- SCHAPIRE R. E. & SINGER Y. (2000). Boostexter : A boosting-based system for text categorization. *Mach. Learn.*, **39**(2-3), 135–168.
- TITOV I. & McDONALD R. (2008). Modeling online reviews with multi-grain topic models. In *Proceedings of the 17th International Conference on World Wide Web, WWW '08*, p. 111–120, New York, NY, USA : ACM.
- TORRES-MORENO J. M., EL-BÈZE M., BÉCHET F. & CAMELIN N. (2007). Comment faire pour que l'opinion forgée à la sortie des urnes soit la bonne ? application au défi deft 2007. *Actes du troisième DÉfi Fouille de Textes*, p. 129.
- VAPNIK V. N. (1995). *The Nature of Statistical Learning Theory*. Springer-Verlag New York, Inc.
- WU Y., ZHANG Q., HUANG X. & WU L. (2009). Phrase dependency parsing for opinion mining. In *Proceedings of EMNLP '09*, p. 1533–1541, Stroudsburg, PA, USA : Association for Computational Linguistics.
- XUEKE X., XUEQI C., SONGBO T., YUE L. & HUAWEI S. (2013). Aspect-level opinion mining of online customer reviews. *Communications, China*, **10**(3), 25–41.