Modèles de langues pour la détection d'opinions dans les blogs

Faiza Belbachir, Mohand Boughanem, Lynda Zaoui

Modèles de langue pour la détection d'opinions dans les blogs

Faiza Belbachir[1,2] **Mohand Boughanem**[1] **Lynda Zaoui**[2]

[1] *Université Toulouse IRIT, UMR 5505 CNRS, France.*
[2] *Université USTO, laboratoire LSSD, département d'informatique, Algérie*

Faiza.Belbachir@irit.fr,Mohand.Boughanem@irit.fr,zaoui_lynda@yahoo.fr

RÉSUMÉ. *Cet article décrit une approche de recherche de documents pertinents vis-à-vis d'une requête et exprimant une opinion. Afin de détecter si un document est porteur d'opinion (i.e. comporte de l'information subjective), nous proposons de le comparer à des sources d'information qui comportent du contenu de type opinion. L'intuition derrière cela est la suivante : un document ayant une similarité forte avec des sources d'opinions, est vraisemblablement porteur d'opinion. Pour mesurer cette similarité, nous exploitons des modèles de langue. Nous modélisons le document et la source (référence) porteuse d'opinions par des modèles de langue, nous évaluons ensuite la similarité de ces modèles. Plusieurs expérimentations ont été réalisées sur des collections issues de TREC. Les résultats obtenus valident notre intuition.*

ABSTRACT. *This article describes an opinion retrieval approach which aims at retrieving relevant and opinionated documents w.r.t. a query. To detect whether a document is opinionated (i.e. contain subjective information), we compare it with opinionated sources that contain subjective information. The intuition is the following, a document having a strong similarity with opinionated sources is likely to be opinionated. To measure this similarity we use language models. We model the document and the source of opinions using language models, we estimate then the similarity of these two models. Several experiments were carried out on TREC collection. The results showed the effectiveness of our approach.*

MOTS-CLÉS : *recherche d'information, blogs, détection d'opinions, modèle de langue.*

KEYWORDS: *information retrieval, blogs, opinions detection, language model.*

1. Introduction

Depuis l'avènement d'Internet, de nombreuses formes de contenu ont été générées par les utilisateurs, y compris les pages personnelles, les discussions et les blogs. Ces derniers sont un moyen facile pour l'expression des avis personnels, le partage des sentiments, ou pour commenter différents sujets. A cause de leur popularité les blogs ont attiré beaucoup d'attention dans les communautés du traitement automatique de la langue naturelle et de la recherche d'information (Adar, Adamic, 2005 ; Agarwal *et al.*, 2008 ; Ding *et al.*, 2008). La présence d'informations de nature subjective apparaît de manière très visible dans les blogs. Les bloggeurs postent des commentaires, font part de leurs sentiments, et diffusent leurs opinions sur divers sujets.

Ces opinions ont une grande importance dans plusieurs domaines (politique, commercial, ou industriel), il serait important alors d'arriver à déterminer les informations objectives des informations subjectives. On pourrait également être amené à rechercher des documents porteurs d'opinions sur un sujet donné. On parle alors de recherche d'opinions, en anglais *opinion retrieval*.

Le challenge dans cette tâche est d'arriver à trouver (sélectionner) des documents qui soient à la fois pertinents pour un sujet donné et porteurs d'opinion sur le sujet. En effet, si la recherche d'information thématique permet de répondre au critère de pertinence, une des problématiques majeure de cette tâche est de répondre au second critère, outre la question relative à l'identification de documents porteurs d'opinions (on parle ainsi de documents subjectifs) ; il faudrait que l'opinion exprimée dans le document porte sur le sujet de la requête. Ceci n'est évidement pas certain, car un document peut traiter différents sujets en même temps.

De manière générale, les approches de recherche d'opinions se basent sur un processus à deux étapes. La première étape s'occupe de rechercher les documents potentiellement pertinents vis-à-vis de la requête. La seconde, quant à elle, consiste à sélectionner parmi eux uniquement ceux porteurs d'opinions.

Les approches relatives à la tâche de détection d'opinion se divisent en deux classes : celles qui exploitent des lexiques d'opinions et celles qui se basent sur l'apprentissage automatique. Certains travaux combinent le lexique et l'apprentissage automatique. Les approches basées sur le lexique utilisent des listes (voire dictionnaire, thésaurus) de mots subjectifs (mots exprimant une opinion). Si un document comporte des mots subjectifs alors il est considéré comme un document de type opinion (Mishne, 2006 ; Oard *et al.*, 2006). Les approches basées sur l'apprentissage automatique utilisent différents classifieurs tels que SVM (Machine à Vecteur de Support) (Cortes, Vapnik, 1995) et Naïve Bayes (Lewis, 1998) entraînés sur des corpus de mots, de phrases ou de documents, annotés comme étant subjectifs (Seki *et al.*, 2007 ; Q. Zhang *et al.*, 2007).

Ces approches dépendent des lexiques ou de la collection d'apprentissage utilisés. Ces ressources peuvent ne pas être disponibles ou non appropriées au langage utilisé dans les documents que l'on recherche. De plus, ils nécessitent un travail préparatoire d'élaboration des listes dans les méthodes basées sur le lexiques ou d'annotation de documents dans le cas de l'apprentissage automatique.

Afin de remédier à ces limites, au lieu d'exploiter des ressources préalablement préparées pour la tâche, nous proposons une approche qui exploite des sources d'informations externes disponibles, ouvertes, comportant effectivement des informations subjectives (des opinions). Plus précisément, nous supposons que si un document est similaire aux documents de la source d'opinions, il est vraisemblablement porteurs d'opinions. Pour estimer cette vraisemblance, nous proposons de modéliser le document à tester et la source d'opinions (nommée également référence dans cet article) par des modèles de langue, et de mesurer la similarité des deux modèles. Plus cette similarité est grande et plus le document est vraisemblablement subjectif.

Le reste de l'article est organisé en comme suit. Dans la première section nous présentons quelques travaux connexes. Nous les divisons en deux catégories, ceux qui exploitent les lexiques et ceux basés sur l'apprentissage automatique. Dans la deuxième section, nous exposons les modèles de langue proposés pour la détection d'opinions. Dans la troisième section, nous présentons les expérimentations ainsi que les résultats obtenus et nous concluons ensuite en listant quelques perspectives.

2. État de l'art

Les approches de recherche d'opinions se divisent en deux classes, celles qui utilisent l'apprentissage automatique et celles qui se basent sur des ressources de mots subjectifs. Nous décrivons dans ce qui suit ces deux classes d'approches.

2.1. Travaux basés sur l'apprentissage automatique

Les approches basées sur l'apprentissage automatique utilisent différents classifieurs entraînés sur des collections de mots, phrases ou documents annotés. Différentes caractéristiques sont alors utilisées pour l'apprentissage.

(Missen *et al.*, 2012) entraînent un classifieur (SVM) en exploitant plusieurs caractéristiques, telles que l'émotivité (correspondant au nombre d'adjectifs, d'adverbes, de verbes et de noms), la subjectivité (mesurée par la moyenne des scores de subjectivité des synsets qui sont les différents sens du mot dans SentiWordNet (*SWN*), la réflexibilité (correspondant aux pronoms qui s'adressent à la personne comme « I, MY, MYSELF» dans le document) et l'adressabilité (formée de pronoms qui s'adressent aux autres personnes comme « YOU, YOURS »). (Wang *et al.*, 2008) entraînent également un classifieur (SVM) en

fonction de la longueur du document, du nombre de mots positifs, négatifs et objectifs.

D'autres travaux (Yang *et al.*, 2006) utilisent une régression logistique entraînée sur différents types de collections. Ils utilisent en particulier, la collection exploitée dans (Pang, Lee, 2004) qui se compose de 5 000 phrases subjectives et de 5 000 phrases objectives provenant des commentaires postés par les *reviewers* du site Amazon.com sur cinq produits électroniques tels que appareil photo, DVD et juke-boxes, un ensemble de 2 041 phrases positives et 2 217 phrases négatives sélectionnées par (Hu, Liu, 2004), et un ensemble de 1 201 phrases positives et 1 240 phrases négatives extraites manuellement de la collection de TREC.

L'avantage de ce type d'approches est l'automatisation complète du processus de recherche de documents exprimant une opinion mais les inconvénients majeurs résident d'une part dans la nécessité d'élaborer préalablement une collection d'apprentissage annotée, et d'autre part, dans la dépendance de ces approches vis-à-vis de la collection d'apprentissage et du classifieur utilisés.

2.2. Travaux basés sur le lexique

Les approches basées sur le lexique exploitent des structures (lexiques, dictionnaires, thésaurus) de mots subjectifs (mots porteurs d'opinions) prédéfinies ou construites automatiquement. Pour évaluer si un document, ou une phrase, exprime une opinion, ces approches mesurent la similarité du contenu du document avec les termes subjectifs. D'autres approches proposent d'étendre la requête de l'utilisateur en y intégrant des mots subjectifs, puis évaluer la similarité requête-document.

Ces approches se distinguent principalement par le type de structures qu'elles exploitent et le traitement effectué au niveau de la requête. Plusieurs travaux se sont basés sur des lexiques construits automatiquement à partir de documents de type opinion. Ces documents peuvent provenir directement de la collection de documents traitée (celle dans laquelle on recherche des opinions) (Santos *et al.*, 2009 ; He *et al.*, 2008 ; Amati *et al.*, 2008 ; Gerani *et al.*, 2009) ou de documents provenant de sources externes (Yang *et al.*, 2006).

Dans le même principe (Na *et al.*, 2009 ; Missen *et al.*, 2013) exploitent des ressources externes, tel que SentiwordNet (*SWN*) (Baccianella *et al.*, 2010). Ces approches pondèrent chaque terme du document en fonction de son degré de subjectivité dans *SWN*. Le score d'opinion du document est calculé en fonction de ces pondérations et de la fréquence des termes dans le document. D'autres travaux combinent plusieurs ressources différentes, en particulier (Yang *et al.*, 2007) exploitent quatre ressources. La première est un lexique construit selon les termes les plus fréquents dans l'ensemble des blogs porteurs d'opinions et les moins fréquents dans des blogs qui n'expriment pas d'opinions. La deuxième

ressource est plus générale, elle se base sur le lexique des termes subjectifs de Wilson (Wilson *et al.*, 2003). Un troisième lexique est plus spécifique au langage des blogs, il comporte les mots rares qui expriment une forte opinion tels que « soo, good ». Le quatrième lexique regroupe tous les pronoms tels que « I » et « you ». Le dernier lexique est construit manuellement, il contient des verbes et des adjectifs de plusieurs ressources lexicales. Le score d'opinion du document est calculé en fonction de la fréquence de ces termes dans le document.

D'autres approches proposent de reformuler la requête en y intégrant des mots subjectifs du lexique. (M. Zhang, Ye, 2008) proposent d'étendre les requêtes à partir des termes du lexique *General Inquiery* en y intégrant les adjectifs qui entourent les termes de la requête.

L'avantage de ces approches est la simplicité de la méthode de détection de document porteur d'opinion. Les limites de ces approches résident tout d'abord dans la disponibilité et l'adéquation des ressources pour la tâche de recherche étudiée. En effet, quand le lexique est interne (extrait de la collection étudiée), il déterminera mieux les mots subjectifs relatifs aux sujets traités dans la collection mais sera spécifique à cette collection. Si le lexique est externe, il sera plus général mais moins approprié à la collection étudiée.

Afin de pallier les limites listées dans les deux sections, nous proposons d'exploiter simultanément des ressources lexicales, en l'occurrence SentiWordNet (SWN), et une ressource générale ouverte, disponible, comportant des textes de type opinion, en l'occurrence la collection IMDb. Cette dernière comporte les avis émis par les internautes sur des films. Ces avis peuvent être considérés comme étant propres aux films, mais nous allons montrer qu'ils peuvent être exploités dans un cadre plus général. Ensuite, afin de prendre en compte conjointement la pertinence et l'opinion, nous proposons un modèle probabiliste unifié basé sur un modèle statistique de langue.

3. Approche proposée

Afin d'évaluer si un document exprime une opinion, nous proposons de mesurer sa similarité avec les documents avérés de type opinion. Pour cela, nous utilisons les modèles génératifs de langue. Plus précisément, nous considérons qu'un document exprime une opinion s'il est généré par un modèle de langue de type opinion. Pour construire ce modèle, nous nous appuyons sur une collection de référence comportant des documents d'opinions. Nous estimons le modèle d'opinion à partir de ces documents. De même pour le document à analyser, nous le modélisons également sous forme d'un modèle de langue.

Pour évaluer le degré de subjectivité (la présence d'opinions dans le document) nous mesurons la similarité entre les deux modèles en utilisant la divergence de Kullback-Leibler (KL_divergence) (Zhai, Lafferty, 2001). Ces deux

modèles peuvent êtres estimés de différentes manières. Nous les détaillons dans les sections qui suivent.

3.1. Modèle de document

Le modèle du document peut être estimé de deux façons, soit indépendamment de la collection de référence ou de manière dépendante de cette collection. Ceci nous permettra d'évaluer si la prise en compte de la collection de référence dans le modèle du document à analyser peut avoir un effet sur la détection d'opinion.

3.1.1. Modèle de langue indépendant

Le modèle document, noté $P_{ML}(w|D)$, indépendant de la collection de référence est estimé par un simple maximum de vraisemblance, en se basant sur la fréquence des termes dans le document (équation 1).

$$ModIndep(D) = P_{ML}(w|D) = \frac{fr(w,d)}{|d|} \tag{1}$$

Où $fr(w,d)$ est la fréquence d'un terme w dans le document d et $|d|$ est la somme des fréquences de tous les termes du document d.

Ce modèle est assez classique. La notion d'opinion n'est pas du tout explicitée. Il modélise juste la distribution des termes dans le document.

3.1.2. Modèle de langue dépendant

Nous introduisons la notion d'opinion dans ce modèle. Une manière intéressante de prendre en compte cette notion est par exemple de renforcer *booster* les mots susceptibles d'exprimer une opinion. Pour ce faire, nous proposons d'estimer le modèle du document en le lissant avec un modèle d'opinion estimé à partir de documents de la collection de référence. Ce lissage permet de renforcer les termes présents dans le document et dans la collection de référence. Le modèle du document est alors défini comme suit, en utilisant le lissage de Jelineck Mercer (Mercer *et al.*, 1983).

$$ModDep(D) = P_{JM}(w|D) = \lambda * P_{ML}(w|D) + (1-\lambda) * P_{ML}(w|R) \tag{2}$$

$P_{ML}(w|R)$ est défini par le maximum de vraisemblance soit :

$$P_{ML}(w|R) = \frac{fr(w,R)}{|R|} \tag{3}$$

Où $fr(w,R)$ représente la fréquence du terme w dans la collection de référence R et $|R|$ représente la somme des fréquences des termes de la collection de référence.

Afin de renforcer davantage cette notion de subjectivité, nous proposons de prendre en compte la subjectivité à priori d'un terme en se basant sur la ressource lexicale (SWN). Dans cette ressource, la subjectivité est mesurée selon trois scores (($Obj(w)$, $Pos(w)$, $Neg(w)$) qui représentent respectivement les scores objectif (quand ce score est élevé le terme est objectif, n'exprime donc pas d'opinion), positif (degré de l'opinion positive) ou négatif (degré de l'opinion négative) d'un synset de SentiWordNet (Dans la terminologie de WordNet un synset est une entrée de WordNet, qui représente un ensemble de mots synonymes). Ces scores sont dans l'intervalle [0, 1] et leur somme pour un synset est égale à 1. Il est à noter qu'un terme donné peut avoir plusieurs sens, et donc appartenir à plusieurs synsets de SWN et avoir des valeurs de subjectivité différentes dans les différents synsets.

Le nombre total de synsets dans lequel un terme apparaît représente le nombre total de sens pour ce terme. Par exemple le « synset » « Estimable », correspondant au sens « peut être calculé ou estimé » de l'adjectif estimable, il a un score objectif de 1,0 et des scores positif et négatif égaux à zéro. Un autre sens pour le même terme est « Dignes de respect ou en haute estime », qui a un score positif de 0,75, un score négatif de 0,0 et un score objectif de 0,25. Donc en cherchant la subjectivité d'un terme dans SWN, une façon simple de la mesurer est de prendre la moyenne de subjectivité (positive et négative) des synsets dans lesquels le terme apparaît. En fait, cette approche est assez simpliste, elle ne ne fait aucune désambiguïsation des termes (ne choisit pas le bon sens du terme considéré). Le lecteur désirant avoir plus de détails sur cette question peut se référer à (Baziz *et al.*, 2005).

Nous calculons ainsi la moyenne du score de subjectivité d'un terme en ajoutant le score positif et le score négatif de tous les sens de ce terme et divisons ensuite le score total par le nombre total des sens du terme (Missen, Boughanem, 2009).

$$Subj(w) = \sum_{si \in sens(w)} \frac{(Neg(si) + Pos(si))}{|sens(w)|} \tag{4}$$

Où $Neg(si)$ est le score négatif du sens si du terme w dans le dictionnaire SWN, $Pos(si)$ est le score positif et $|sens(w)|$ est le nombre de sens du terme retrouvé dans SWN.
En intégrant cette notion de subjectivité, le modèle du document $P(w|D)$ sera alors donné par l'équation qui suit :

$$ModDepSubj(D) = P_{JM\_Sub}(w|D) = Subj(w) * P_{JM}(w|D) = \lambda * P_{ML}(w|D) * Subj(w) + (1 - \lambda) * Subj(w) * P_{ML}(w|R) \tag{5}$$

3.2. Modèle de référence

Le modèle de référence $P(w|R)$ est également estimé de deux manières. La première combine le modèle de la collection de référence (R) avec le modèle de la collection à analyser (C). Cela permet de favoriser les termes qui appartiennent à la collection à analyser, et de ne pas obtenir une probabilité nulle si le terme du document ne se trouve pas dans la collection de référence. Le modèle est alors estimé comme suit :

$$ModDep(R) = P_{JM}(w|R) = \lambda * P_{ML}(w|R) + (1 - \lambda) * P_{ML}(w|C) \quad (6)$$

Et $P_{ML}(w|C)$ est représenté comme suit

$$P_{ML}(w|C) = \frac{fr(w, C)}{|C|} \quad (7)$$

Où $fr(w, C)$ représente la fréquence du terme w dans la collection d'analyse C et $|C|$ est la taille de cette collection.

En ce qui concerne la seconde manière de la même façons que le modèle de document, nous proposons de favoriser les mots subjectifs en y intégrant leur score de subjectivité, dans le modèle de langue de référence soit :

$$ModDepSubj(R) = P_{JM\_Sub}(w|R) = Subj(w) * P_{JM}(w|R) = \lambda * P_{ML}(w|R) * Subj(w) + (1 - \lambda) * P_{ML}(w|C) * Subj(w) \quad (8)$$

3.3. Score d'opinion

Afin d'évaluer le score d'opinion d'un document, nous adaptons le calcul de score initialement utilisé dans le domaine de la recherche d'information (calcul de score de pertinence) pour la détection d'opinion. Pour ce faire, nous mesurons la similarité entre son modèle de langue et celui de l'opinion. Comme nous l'avons déjà mentionné, nous utilisons pour cela une fonction classique, la divergence de Kullback-Leibler (Zhai, Lafferty, 2001). Ce score est représenté par la formule suivante :

$$Score\_KL\_R(D, R) = \sum_{w \in D} P(w|D) * log \frac{P(w|D)}{P(w|R)} \quad (9)$$

Où $P(w|D)$ et $P(w|R)$ sont les modèles de langue respectivement, du document et de la collection d' opinions. Cette fonction mesure en fait plutôt la divergence entre les distributions de probabilités. Plus le score est faible plus le document est similaire à la collection d'opinions, donc est vraisemblablement porteur d'opinion.

Nous exploitons également une version assez simpliste, qui mesure ce score comme une probabilité jointe de tous les termes du document, soit :

$$Score\_Prod\_R(D) = \prod_{w \in D} P(w|D) \tag{10}$$

L'intuition ici est la suivante : la distribution des termes dans le document est censée modéliser l'importance des ces termes dans le document, comme l'importance des termes de type opinion a été boostée, nous pensons que plus le score $Score\_Prod\_R(D)$ est élevé plus le document a des chances de contenir davantage d'opinions.

Remarque. – tous les scores calculés dans cette section ne font pas du tout référence à la notion de pertinence, qui est au coeur de la recherche d'information. Rappelons que notre but est d'étudier uniquement la dimension opinion. En ce qui concerne la dimension pertinence, nous avons considéré qu'elle est calculée par ailleurs, en utilisant n'importe quel modèle de recherche d'information. De ce fait, pour renvoyer la liste de documents répondant à la requête et exprimant une opinion, nous proposons juste de faire une combinaison linéaire entre le score de pertinence et le score d'opinion. Ce point est discuté dans la dernière partie de la section expérimentations.

4. Expérimentations

Les expérimentations sont réalisées sur deux collections, l'une de TREC Blog Track 2006 (Macdonald, Ounis, TREC 2006) qui représente la collection à analyser et la seconde est IMDb (Internet Movie Database) [1] qui représente la collection de référence.

La collection TREC Blog Track 2006 comporte plus de 3,2 millions de post blogs extraits durant une période de 11 semaines de Décembre 2005 à Février 2006. TREC propose chaque année un ensemble de 50 topics (un topic correspond à une requête et il est appelé aussi sujet) ainsi que leurs jugements de pertinence. Les blogs sont annotés par des spécialistes du domaine de la manière suivante : 0 pour les blogs non pertinents, 1 pour les blogs pertinents, 2 pour les blogs à opinion négative, 3 pour ceux à opinion mixte et 4 pour ceux à opinion positive. Cette collection sera considérée comme la base de vérité et elle est nommée QRELS.

Concernant la collection IMDb, elle représente une base de données en ligne sur le cinéma, la télévision et les jeux vidéo. Toute personne peut poster et partager des avis sur n'importe quel film. Le site a été créé le 17 Octobre

1. `http://www.cs.cornell.edu/people/pabo/movie-review-data/`.

1990 et est devenu parmi les sites les plus visités au monde (classé au rang 38 dans le monde) et a plus de 57 millions de visiteurs par mois. L'intérêt de cette collection est qu'elle contient un grand nombre d'opinions et d'avis. Les auteurs (Pang, Lee, 2004) se sont intéressés à cette collection et en ont extrait un ensemble de documents (2 000 documents) contenant des opinions que nous utilisons pour nos expérimentations.

4.1. Résultats

Pour évaluer notre approche, nous procédons de la manière suivante. Nous prenons les 1 000 premiers documents pour chaque topic (Baseline 4 de TREC). Cette liste est en fait fournie par les évaluateurs de TREC. Elle correspond aux 1 000 premiers documents répondant thématiquement à la requête. TREC fournit cette liste afin que les participants à TREC partent du même sous-ensemble de documents pertinents, par conséquent l'impact de l'approche d'opinion proposée peut être évalué indépendamment de la performance du modèle de pertinence thématique utilisé.

Notre objectif est alors de réordonner ces documents en fonction de leur opinion. Les résultats des évaluations sont présentés en termes de précision moyenne (*Average Precision* (AP)) quand il s'agit d'une requête, ou de la moyenne des AP (*Mean Average Precision* (MAP)) pour un ensemble de requêtes et de la précision à 10 documents notée (P@10).

Nous avons réalisé quelques expérimentations préliminaires qui nous ont permis de fixer le paramètre λ égal à 0,6. Nous avons évalué l'impact de différentes configurations :

1. L'impact de la subjectivité à priori calculée selon SentiWordNet,
2. L'impact de la collection de référence dans le modèle du document,
3. L'impact de la fonction de calcul du score d'opinion,
4. la performance du score final combinant le score d'opinion et le score de pertinence.

4.1.1. Impact de la subjectivité selon SentiWordNet

Dans cette expérimentation le but est de mesurer l'impact du facteur subjectivité calculé selon *SWN* sur la détection d'opinion. Pour ce faire nous comparons les résultats obtenus en prenant les deux modèles (document, référence) selon qu'ils prennent en compte la subjectivité selon SWN ($ModDepSubj(D)$, $ModDepSubj(R)$) ou non ($ModDep(D)$, $ModDep(R)$). Le score d'opinion est donné par $Score\_KL\_R(D, R)$(formule 9).

Le tableau 1 présente les valeurs de MAP et P@10 obtenues pour l'ensemble des 50 requêtes. Les résultats montrent que la différence des performances entre les deux configurations est très faible, et non statistiquement significative. On

pourrait donc conclure que l'utilisation de SentiWordNet n'a pas d'impact dans les modèles proposés.

Tableau 1. Les résultats des mesures MAP et P@10 pour le modèle qui prend en compte la subjectivité selon SWN *et celui qui n'en tient pas compte*

Configuration	MAP	P@10
Sans Subjectivité	0,1690	0,3271
Avec Subjectivité	0,1694	0,3273

Quand on observe les requêtes de manière individuelle tel qu' illustré dans la figure 1 où *les topics* de TREC 2006 sont représentés en abscisse et la précision moyenne pour chaque topic en ordonnée, on constate que les performances entre les différentes représentations sont similaires. (AP (*sans_sub*) représente les résultats pour les topics qui ne prennent pas en compte la subjectivité et AP (*avec_sub*) ceux qui prennent en compte la subjectivité).

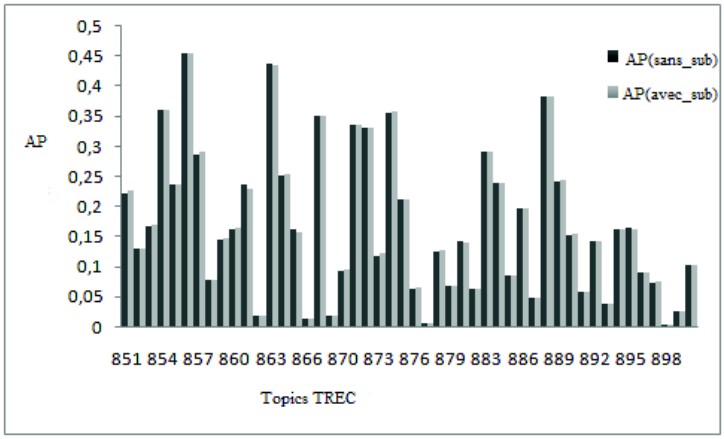

Figure 1. Impact de la subjectivité

D'après les résultats obtenus, il en ressort que le lexique SWN n'a pas d'impact sur la détection d'opinion dans notre approche. Ceci peut s'expliquer soit par le fait que les distributions des termes en fonction de leur subjectivité (degré d'opinion) sont les mêmes dans SWN et dans la collection de référence, ce qui nous semble invraisemblable. Soit par le fait, le plus plausible, que peu de termes de type opinion sont effectivement présents dans SWN. Il est à noter que dans nos expérimentations, si un terme n'est pas présent dans SWN, nous ne considérons pas son poids selon SWN, mais on garde son poids d'origine

$P_{JM}(w|R)$ et $P_{JM}(w|D)$. Par conséquent, dans ce cas les modèles avec et sans SWN sont identiques. Nous devrons approfondir cette étude dans nos futures investigations.

4.1.2. Impact de la collection de référence

Dans cette expérimentation, nous comparons le modèle du document qui prend en considération la collection de référence ($ModDepSubj(D)$) avec sa version sans référence ($ModIndep(D)$). Quant au modèle de référence, il sera basé sur sa représentation optimale suivante ($ModDep(R)$).

Tableau 2. MAP et P@10 pour le modèle de document lissé avec la collection de référence vs. le modèle non lissé

MÉTHODE	MAP	P@10
Avec Référence	0,1690 (48%)	0,3271 (133%)
Sans Référence	0,1136	0,1402

Le tableau 2 montre les résultats obtenus en terme de MAP et de P@10 sur l'ensemble des requêtes. Nous remarquons clairement que la prise en compte de la collection de référence permet une amélioration très significative des résultats soient 48 % d'amélioration au niveau de MAP et 133 % au niveau de P@10 comparativement aux modèles qui n'utilisent pas la collection de référence.

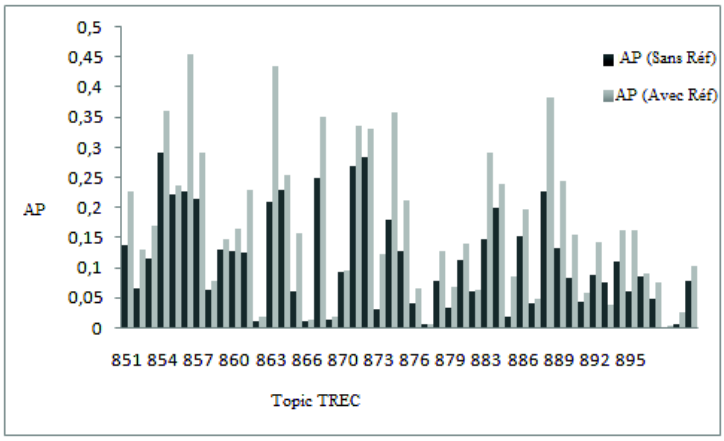

Figure 2. Impact de la collection à référence

L'analyse des résultats requête par requête présentée sur la figure 2, montre que le modèle lissé par la collection de référence présente des performances plus élevées pour 48 topics sur 50. AP(sans Réf) représente les résultats pour les

topics qui ne prennent pas en compte la collection de référence et AP(avec Réf) pour ceux qui la prennent en compte.

Nous pouvons ainsi conclure que l'exploitation directe d'une ressource externe porteuse d'opinions permet en effet d'identifier les documents porteurs d'opinions.

4.1.3. Impact de la fonction de score d'opinion

Nous comparons ici les deux fonctions de score proposées à savoir la divergence de Kullback-Leibler $Score\_KL\_R(D)$ et le $Score\_Prod\_R(D)$. Nous avons pris pour les modèles de document et de la collection de référence les représentations optimales suivantes (ModDepSubj(D), $ModDepSubj(R)$).

Tableau 3. Les résultats des mesures MAP et P@10 pour le modèle qui se base sur un score produit ou sur une similarité pour ré-ordonner les documents à opinions

TOPIC	MAP	P@10
Prod	0,1063	0,1187
KL-divergence	0,1690	0,3271

Le tableau 3 liste les résultats de ces deux configurations. La conclusion est claire, et d'ailleurs attendue. La KL-divergence donne des résultats largement supérieurs à ceux obtenus par un simple produit.

Afin de mieux comprendre le comportement de ces fonctions de score, nous avons analysé la distribution des scores $score\_KL\_R(D, R)$ et $score\_prod\_R(D)$ dans les documents qui contiennent des opinions et dans les documents qui ne contiennent pas d'opinions.

Les résultats sont représentés dans les figures 3 et 4. En abscisses, les 50 topics de TREC 2006 et en ordonnées, respectivement la différence entre la moyenne des scores des documents pertinents (à opinion) et la moyenne des documents non opinion. Seuls les documents effectivement jugés par TREC sont utilisés.

Les résultats de la figure 3, listant la différence des scores dans le cas de la fonction $score\_Prod\_R(D)$, sont beaucoup plus mitigés. Il y a la moitié des topics (soit 25 sur 50 topics) pour lesquels les scores diffèrent, et pour les 25 autres topics il n'y pas pas de différence. La figure 4 quand a elle, montre qu'il y a une différence claire entre les scores de documents porteurs d'opinions et ceux de type non opinions (soit plus de 42 topics sur 50 affichent cette différence).

Nous constatons que la KL-divergence donne des résultats largement supérieurs à ceux obtenus par le $Score\_Prod\_R(D)$. Ce qui s'explique par le fait que la notion d'opinion est explicitement insérée et de deux manières dans la KL divergence, elle est exprimée dans la probabilité $P(w|D)$ et dans la proba-

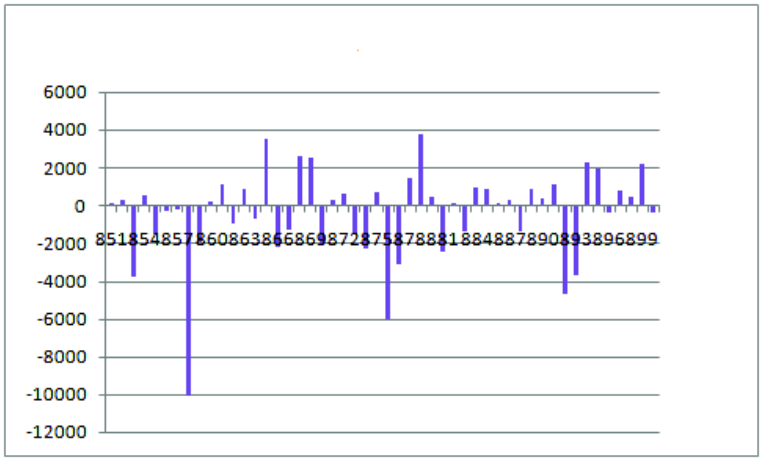

Figure 3. La différence des scores moyens des documents opinion et non-opinion calculés selon score_prod_R(D)

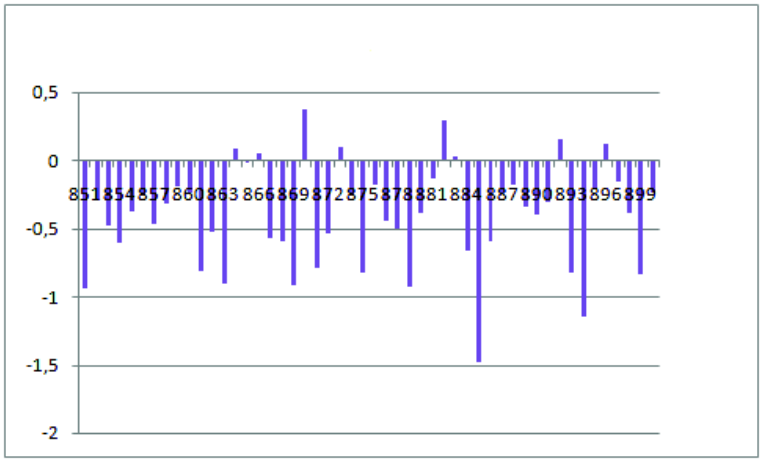

Figure 4. La différence des scores moyens des documents opinion et non-opinion calculés selon score_KL_R(D)

bilité $P(w|R)$. Tandis que pour le score qui se base sur le produit, la notion d'opinion est exprimée uniquement dans la probabilité $P(w|D)$ ce qui explique sa faible performance.

4.1.4. Évaluation du modèle combinant pertinence et opinion

Tous les résultats que nous avons décrits précédemment ne prennent pas en compte la pertinence des documents. Notre but étant d'évaluer uniquement la dimension opinion dans le processus de recherche d'opinion. Comme la tâche

qui nous intéresse est la recherche de documents pertinents et exprimant une opinion sur un sujet (une requête), et afin de mieux répondre à cette tâche, nous proposons donc de combiner le score d'opinion d'un document avec son score de pertinence.

Le score d'opinion d'un document est calculé selon $score\_KL\_R(D)$ basé sur les représentations optimales suivantes (ModDepSubj(D), $ModDepSubj(R)$). Le score de pertinence d'un document est donné par la baseline de TREC. Nous avons utilisé une simple combinaison linéaire soit:

$$Score\_Final(D) = \alpha Score\_Pertinence(D, q) + (1 - \alpha) Score\_Opinion(D) \quad (11)$$

Avec q qui représente la requête, D le document et α un paramètre de lissage, Nous avons réalisé quelques expérimentations préliminaires, qui ont conduit a fixer le paramètre à 0,4.

Le tableau 4 liste les résultats du $Score\_Final(D)$ avec la meilleure baseline de TREC (Baseline 4). Les résultats du tableau 4 montrent que la méthode qui

Tableau 4. Les résultats des mesures MAP et P@10

TOPIC	MAP	P@10
Score_Final	0,3063	0,5542
Baseline 4 TREC	0,3022	0,5240

se base sur le $Score\_Final$ (MAP égale à 0,3063 et P@10 égale à 0,5542) est meilleure que la Baseline 4 de TREC, soit plus de 2 % d'amélioration au niveau de MAP et de 5 % au niveau de P@10.

Il en ressort donc que l'exploitation directe d'une ressource ouverte et disponible de type opinion, la collection IMDb dans notre cas peut être utilisée dans un cadre plus général, en l'occurrence dans la collection de TREC Blogs. Ce qui nous semblait invraisemblable au début, car cette ressource comporte des avis des internautes sur des films et devrait être spécifique uniquement à ce domaine.

5. Conclusion

Cet article aborde la question de la détection d'opinions. Nous partons du fait qu'un document (un blog dans notre cas) contient des opinions s'il est similaire à une source d'information de type opinion. Pour ce faire, nous avons modélisé le document et la collection de référence (opinions) par des modèles statistiques de langue, puis nous proposons de mesurer la similarité entre ces modèles pour déterminer si le document est subjectif. Les différentes expérimentations que nous avons menées ont montré que notre hypothèse sur l'exploitation de sources d'opinions sans analyse préalable (extraction de mots subjectifs) est viable. Nous avons effectivement amélioré de manière significative

nos résultats comparativement à la baseline donnée par TREC et également par rapport aux différentes configurations que nous avons considérées.

Nos travaux futurs vont se concentrer sur deux points. Le premier concerne une meilleure modélisation de la source d'opinions. En fait, dans notre approche tous les termes de la collection de référence sont utilisés et il serait intéressant de trouver une manière qui permet de *booster* les termes vraisemblablement subjectifs. Le second point concerne la détection de la polarité de l'opinion. Notre but est d'étendre le modèle de langue pour identifier si l'opinion exprimée dans le document est positive, négative ou neutre.

Bibliographie

Adar E., Adamic L. A. (2005). Tracking information epidemics in blogspace. In *Proceedings of the 2005 ieee/wic/acm international conference on web intelligence*, p. 207–214. Washington, DC, USA, IEEE Computer Society. Consulté sur http://dx.doi.org/10.1109/WI.2005.151

Agarwal N., Liu H., Tang L., Yu P. S. (2008). Identifying the influential bloggers in a community. In *Proceedings of the 2008 international conference on web search and data mining*, p. 207–218. New York, NY, USA, ACM. Consulté sur http://doi.acm.org/10.1145/1341531.1341559

Amati G., Ambrosi E., Bianchi M., Gaibisso C., Gambosi G. (2008). Automatic construction of an opinion-term vocabulary for ad hoc retrieval. In *Proceedings of the ir research, 30th european conference on advances in information retrieval*, p. 89–100. Berlin, Heidelberg, Springer-Verlag. Consulté sur http://dl.acm.org/citation.cfm?id=1793274.1793289

Baccianella S., Esuli A., Sebastiani F. (2010, may). Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In N. C. C. Chair) *et al.* (Eds.), *Proceedings of the seventh international conference on language resources and evaluation (lrec'10).* Valletta, Malta, European Language Resources Association (ELRA).

Baziz M., Boughanem M., Aussenac-Gilles N. (2005). Evaluating a conceptual indexing method by utilizing wordnet. In *Clef*, p. 238-246.

Chesley P. (2006). Using verbs and adjectives to automatically classify blog sentiment. In *In proceedings of aaai-caaw-06, the spring symposia on computational approaches*, p. 27–29.

Clark S. W. D. H. M., Beresi U. C. (2008). Rgu at the trec blog track. In *Text retrieval conference.*

Cortes C., Vapnik V. (1995, septembre). Support-vector networks. In, vol. 20, p. 273-297. Hingham, MA, USA, Kluwer Academic Publishers. Consulté sur http://dx.doi.org/10.1023/A:1022627411411

Cronen-Townsend S., Zhou Y., Croft W. B. (2002a). Predicting query performance. In *Proceedings of the 25th annual international acm sigir conference on research and development in information retrieval*, p. 299–306. New York, NY, USA, ACM. Consulté sur http://doi.acm.org/10.1145/564376.564429

Cronen-Townsend S., Zhou Y., Croft W. B. (2002b). Predicting query performance. In *Proceedings of the 25th annual international acm sigir conference on research and development in information retrieval*, p. 299-306. New York, NY, USA, ACM. Consulté sur http://doi.acm.org/10.1145/564376.564429

Ding X., Liu B., Yu P. S. (2008). A holistic lexicon-based approach to opinion mining. In *Proceedings of the 2008 international conference on web search and data mining*, p. 231–240. New York, NY, USA, ACM. Consulté sur http://doi.acm.org/10.1145/1341531.1341561

Drezner D. W., Farrell H. (134(1):15-30, January 2008). The power and politics of blogs. Consulté sur http://www.cs.duke.edu/courses/spring05/cps182s/readings/blogpowerpolitics.pdf

Ernsting B., Weerkamp W., Rijke M. de. (2007). Language modeling approaches to blog postand feed finding. In *Trec.*

Esuli A., Sebastiani F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. In *In proceedings of the 5th conference on language resources and evaluation lrec 06*, p. 417–422.

Gerani S., Carman M. J., Crestani F. (2009). Investigating learning approaches for blog post opinion retrieval. In M. Boughanem, C. Berrut, J. Mothe, C. SoulÃ©-Dupuy (Eds.), *Ecir*, vol. 5478, p. 313-324. Springer. Consulté sur http://dblp.uni-trier.de/db/conf/ecir/ecir2009.html#GeraniCC09

He B., Macdonald C., He J., Ounis I. (2008). An effective statistical approach to blog post opinion retrieval. In J. G. Shanahan *et al.* (Eds.), *Cikm*, p. 1063-1072. ACM. Consulté sur http://dblp.uni-trier.de/db/conf/cikm/cikm2008.html#HeMHO08

Hoang L., Lee S.-W., Hong G., Lee J.-Y., Rim H.-C. (2008). A hybrid method for opinion finding task (kunlp at trec 2008 blog track). In *Trec.*

Hu M., Liu B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth acm sigkdd international conference on knowledge discovery and data mining*, p. 168–177. New York, NY, USA, ACM. Consulté sur http://doi.acm.org/10.1145/1014052.1014073

Jelinek F., Mercer R. L. (1980). Interpolated estimation of Markov source parameters from sparse data. In E. S. Gelsema, L. N. Kanal (Eds.), *Proceedings, workshop on pattern recognition in practice*, p. 381–397. Amsterdam, North Holland.

Jia L., Yu C. T., Zhang W. (2008). Uic at trec 208 blog track. In *Trec.*

Jones K. S., Walker S., Robertson S. E. (2000). A probabilistic model of information retrieval: development and comparative experiments. In *Information processing and management*, p. 779–840.

Kullback S., Leibler R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, vol. 22, p. 49–86.

Lafferty J., Zhai C. (2001). Document language models, query models, and risk minimization for information retrieval. In *Proceedings of sigir*, p. 111-119. USA, ACM. Consulté sur http://doi.acm.org/10.1145/383952.383970

Lee Y., Na S. hoon, Kim J., Nam S. hyob, Jung H. young, Lee J. hyeok. (2008). Kle at trec 2008 blog track: Blog post and feed retrieval. In *In proceedings of trec-08.*

Lewis D. D. (1998). Naive (bayes) at forty: The independence assumption in information retrieval. In *Proceedings of the 10th european conference on machine learning*, p. 4-15. London, UK, UK, Springer-Verlag. Consulté sur http://dl.acm.org/citation.cfm?id=645326.649711

Liao X., Cao D., Tan S., Liu Y., Ding G., Cheng X. (2006). Combining language model with sentiment analysis for opinion retrieval of blog-post. In E. M. Voorhees, L. P. Buckland (Eds.), *Proceedings of trec 2006, gaithersburg*, vol. Special Publication 500-272. (NIST).

Macdonald C., Ounis I. (TREC 2006). *The TREC Blogs06 collection: creating and analysing a blog test collection.*

Manning C. D., Schütze H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA, USA, MIT Press.

Mercer R. L., Bahl L. R., Jelinek F. (1983). A maximum likelihood approach to continuous speech recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 5, n° 2, p. 179-190. Consulté sur http://dblp.uni-trier.de/db/journals/pami/pami5.html#BahlJM83

Mishne G. (2006). Multiple ranking strategies for opinion retrieval in blogs - the university of amsterdam at the 2006 trec blog track. In *Trec.*

Missen M. M. S., Belbachir F., Cabanac G. (2012, septembre). Combining document-level topic dependent and topic independent evidences for opinion retrieval. *Information - Interaction - Intelligence*, vol. 12, n° 1, p. 53–74. Consulté sur http://www.irit.fr/journal-i3/volume12/numero01/article_12_01_04.pdf (SIGRI)

Missen M. M. S., Boughanem M. (2009). Using wordnet's semantic relations for opinion detection in blogs. In *Ecir*, p. 729-733.

Missen M. M. S., Boughanem M., Cabanac G. (2013, mars). Opinion mining: Reviewed from word to document level. *Social Network Analysis and Mining*, vol. 3, n° 1, p. 107–125. Consulté sur http://dx.doi.org/10.1007/s13278-012-0057-9

Montague M., Aslam J. A. (2002). Condorcet fusion for improved retrieval. In *Proceedings of the eleventh international conference on information and knowledge management*, p. 538–548. New York, NY, USA, ACM. Consulté sur http://doi.acm.org/10.1145/584792.584881

Na S.-H., Lee Y., Nam S.-H., Lee J.-H. (2009). Improving opinion retrieval based on query-specific sentiment lexicon. In *Advances in information retrieval*, vol. 5478, p. 734–738. Springer Berlin / Heidelberg. Consulté sur http://www.springerlink.com/content/1706450803x5w354/

Oard D. W., Elsayed T., Wang J., Wu Y., Zhang P., Abels E. G. *et al.* (2006). Trec 2006 at maryland: Blog, enterprise, legal and qa tracks. In *Trec.*

Osman D. J., Yearwood J., Vamplew P. (2007). Using corpus analysis to inform research into opinion detection in blogs. In P. Christen, P. J. Kennedy, J. Li, I. Kolyshkina, G. J. Williams (Eds.), *Data mining and analytics 2007, proceedings of the sixth australasian data mining conference (ausdm 2007), gold coast, queensland, australia, december 3-4, 2007, proceedings*, vol. 70, p. 65-75. Australian Computer Society.

Pang B., Lee L. (2004, July). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd meeting of the association for computational linguistics (acl'04), main volume*, p. 271–278. Barcelona, Spain. Consulté sur http://www.aclweb.org/anthology/P04-1035

Ponte J. M., Croft W. B. (1998). A language modeling approach to information retrieval. In *Proceedings of the 21st annual international acm sigir conference on research and development in information retrieval*, p. 275–281. New York, NY, USA, ACM. Consulté sur http://doi.acm.org/10.1145/290941.291008

Santos R. L. T., He B., Macdonald C., Ounis I. (2009). Integrating proximity to subjective sentences for blog opinion retrieval. In M. Boughanem, C. Berrut, J. Moth, C. SoulÃ©-Dupuy (Eds.), *Ecir*, vol. 5478, p. 325-336. Springer. Consulté sur http://dblp.uni-trier.de/db/conf/ecir/ecir2009.html#SantosHMO09

Seki K., Kino Y., Sato S., Uehara K. (2007). Trec 2007 blog track experiments at kobe university. In *Trec.*

Wang J., Sun Y., Mukhtar O., Srihari R. K. (2008). Trec 2008 at the university at buffalo: Legal and blog track. In E. M. Voorhees, L. P. Buckland (Eds.), *Trec*, vol. Special Publication 500-277. National Institute of Standards and Technology (NIST). Consulté sur http://dblp.uni-trier.de/db/conf/trec/trec2008.html#WangSMS08

Wilson T., Hoffmann P., Somasundaran S., Kessler J., Wiebe J., Choi Y. *et al.* (2005). Opinionfinder: a system for subjectivity analysis. In *Proceedings of hlt/emnlp*, p. 34-35. USA, Association for Computational Linguistics. Consulté sur http://dx.doi.org/10.3115/1225733.1225751

Wilson T., Pierce D. R., Wiebe J. (2003). Identifying opinionated sentences. In *Proceedings of the 2003 conference of the north american chapter of the association for computational linguistics on human language technology: Demonstrations - volume 4*, p. 33–34. Stroudsburg, PA, USA, Association for Computational Linguistics. Consulté sur http://dx.doi.org/10.3115/1073427.1073444

Yang, Si L., Callan J. (2006). Knowledge transfer and opinion detection in the TREC2006 blog track. In *Proceedings of trec.*

Yang, Yu N., Zhang H. (2007). Widit in trec 2007 blog track: Combining lexicon-based methods to detect opinionated blogs. In *Trec.*

Yue Y., Finley T., Radlinski F., Joachims T. (2007). A support vector method for optimizing average precision. In *Proceedings of the 30th annual international acm sigir conference on research and development in information retrieval*, p. 271–278. New York, NY, USA, ACM. Consulté sur http://doi.acm.org/10.1145/1277741.1277790

Zhai C., Lafferty J. (2001). Document language models, query models, and risk minimization for information retrieval. In *Proceedings of the 24th annual international acm sigir conference on research and development in information retrieval*, p. 111–119. New York, NY, USA, ACM. Consulté sur http://doi.acm.org/10.1145/383952.383970

Zhang M., Ye X. (2008). A generation model to unify topic relevance and lexicon-based sentiment for opinion retrieval. In *Proceedings of the 31st annual international acm sigir conference on research and development in information retrieval*, p.

411–418. New York, NY, USA, ACM. Consulté sur http://doi.acm.org/10.1145/1390334.1390405

Zhang Q., Wang B., Wu L., Huang X. (2007). Fdu at trec 2007: Opinion retrieval of blog track. In E. M. Voorhees, L. P. Buckland (Eds.), *Proceedings of the sixteenth text retrieval conference, trec 2007, gaithersburg, maryland, usa, november 5-9, 2007*, vol. Special Publication 500-274. National Institute of Standards and Technology (NIST).

Zhou G., Joshi H., Bayrak C. (2007). Topic categorization for relevancy and opinion detection. In *Trec*.

Zhou L., Twitchell D. P., Qin T., Burgoon J. K., Nunamaker J. F., Jr. (2003). An exploratory study into deception detection in text-based computer-mediated communication. In *Proceedings of the 36th annual hawaii international conference on system sciences (hicss'03) - track1 - volume 1*, p. 44.2–. Washington, DC, USA, IEEE Computer Society. Consulté sur http://dl.acm.org/citation.cfm?id=820748.821356