

EXPÉRIMENTATION D'APPROCHES POUR LA DÉTECTION D'OPINIONS ET DE LEUR POLARITÉ DANS LES BLOGS

Faiza BELBACHIR, Malik Muhammad Saad MISSEN, Guillaume CABANAC, Mohand BOUGHANEM
IRIT/SIG – UMR 5505 CNRS — Université Toulouse 3 – 118 route de Narbonne – F-31062 Toulouse cedex 9

Mots clefs :

Recherche d'information, blogs, détection d'opinion, polarité, classification de requête.

Keywords:

Information retrieval, blogs, opinion detection, polarity, query classification.

Résumé

Cet article décrit notre contribution sur la détection d'opinions et de la polarité dans les blogs. L'approche proposée pour la détection de l'opinion est indépendante des topics et est basée sur le lexique SentiWordNet et l'apprentissage machine. Concernant la polarité, nous avons classifié les topics de TREC 2006 et 2007, et nous avons étudié l'impact positif de cette classification sur la détection de la polarité.

Abstract

This paper describes our work on the tasks of opinion detection and opinion polarity detection in blogs. We propose and evaluate a set of topic independent opinion finding evidences for opinion detection task. The approach we adapt is a combination of lexicon-based and machine-learning approaches. In addition, our work includes the classification of topics of TREC Blog collection for years 2006 and 2007. Finally, we analyze the positive impacts of topic classification on the tasks of opinion detection and opinion polarity detection.

1 Introduction et motivations

De nos jours plusieurs applications et plateformes sur le web nous permettent de déposer des avis, de partager des sentiments et des opinions sur une variété de sujets. Vue l'importance de ces informations dans plusieurs domaines (politique, commercial ou individuel), il serait intéressant de déterminer l'information subjective contenue dans les textes. Plusieurs travaux se sont intéressés à la détection d'opinions et à la détection de la polarité. La détection d'opinions est une tâche qui permet d'extraire les opinions d'un ensemble de documents pertinents pour un sujet donné. Elle est confrontée à des problèmes qui la distinguent de la recherche d'information traditionnelle thématique dont les sujets sont souvent identifiés par des mots-clés seulement. L'opinion peut être exprimée de manière très variée et subtile et donc il est souvent difficile de la déterminer exactement. La classification du sentiment (polarité) est une sous-tâche de la détection d'opinions. Elle consiste de façon générale à déterminer si l'opinion du document sur le sujet est positive ou négative. De ce fait, plusieurs travaux de recherche se sont intéressés à ce problème [4, 16]. Afin d'évaluer les résultats des chercheurs dans le domaine, plusieurs campagnes d'évaluations ont vu le jour. Ces campagnes proposent une collection de test qui comporte un ensemble de documents, un ensemble de topics (sujet) et des jugements de pertinence « relevance judgement » nommés « qrels ». Ces derniers représentent la perception humaine. Les résultats des travaux sont donc comparés aux « qrels » à travers de mesures [1] telles que le rappel, la précision, la MAP, etc.

Dans cet article on s'est intéressé au problème de la détection d'opinions et de la polarité dans les blogs. Le blog est un site Web constitué de billets (Blog post) écrit par une personne selon un rythme périodique. Pour évaluer nos résultats, nous utilisons la collection de test « Blog06 » proposée par la campagne d'évaluation TREC. Pour la détection d'opinions, nous proposons une approche utilisant le lexique SentiWordNet et l'apprentissage machine. Les features (caractéristiques) que nous avons jugées importantes sont l'émotivité, la subjectivité, la réflexivité et l'adressage et sont indépendantes des topics. Une étude comparative a été effectuée pour trouver la combinaison des features permettant la meilleure détection de l'opinion dans les blogs à l'aide du classifieur Regression Logistique. Cette combinaison a été étudiée sur d'autres classifieurs (SVM et Naive Bayes). Afin de prendre en considération le contexte (un mot peut avoir une polarité différente selon le contexte), une catégorisation des topics a été effectuée. Une étude de l'impact de celle-ci sur les performances de la détection de la polarité a été faite.

2 Campagne d'évaluation : TREC Blog Track

Le mot TREC signifie « Text Retrieval Conference » et désigne l'ensemble des conférences organisées par le NIST « National Institute of Standard and Technology » sur la recherche d'information. Plusieurs tâches ont fait l'objet de recherches dans ces conférences, dont le Blog Track qui a été introduit en 2006. Chaque année, de nouvelles tâches sont définies dans la détection d'opinions et la détection de la polarité. Plusieurs universités et laboratoires de recherche participent à TREC et utilisent la collection de test proposée par TREC pour l'évaluation de leurs résultats.

2.1 Collection de test

Une collection de test comporte trois parties [9] un ensemble de documents, un ensemble de topics et les jugements de pertinence « relevance judgements ». Le résultat d'une exécution d'un système de recherche exécutant une tâche sur une collection de test est appelé un « run ».

1. Corpus documentaire

L'ensemble des documents d'une collection de test doit refléter la réalité : diversité des sujets, des formats et des styles. Dans le blog track, deux collections ont été créées par l'université de Glasgow. La première dite collection « Blog 06 » a été utilisée dans TREC 2006, 2007 et 2008. C'est un échantillon des blogs de la blogosphère. Les documents ont été collectés pendant 11 semaines du 6 décembre 2005 au 21 février 2006.

Cette collection est de 148 Go et se compose de 38,6 Go de feed (blogs), de 88,8 Go de permalinks (un billet de blog simple avec tous les commentaires associés) et 20,8 Go de home pages (page d'accueil d'un blog). La deuxième collection dite « Blog 08 » [10] a été créée en 2008 et a été utilisée dans TREC 2009. Elle est plus volumineuse que la précédente et a été collectée de janvier 2008 à février 2009. La taille de cette collection est de 2309 Go et contient 808 feeds, 1445 permalink et 56 homes pages. Chaque document est caractérisé par un unique identifiant appelé DOCNO.

2. Besoins en information (*topics*)

Les topics sont généralement construits spécialement pour la tâche pour laquelle ils sont utilisés. Un topic peut être une entité nommée comme un nom d'une personne, un emplacement ou un organisme mais il peut être aussi un concept, comme un type de technologie, par exemple. Un topic est caractérisé par un identifiant, un titre, une description (généralement une phrase décrivant le topic) et une narration qui décrit l'information que l'individu recherche et qu'il considère comme pertinente. La campagne d'évaluation propose chaque année de nouveaux topics, 50 topics ont été définis en 2006, 50 autres en 2007 et 50 topics en 2008.

3. Jugements de pertinence (*qrels*)

Les jugements de pertinence ne peuvent pas être faits manuellement pour deux raisons majeures. La première est que la collection de documents est trop grande et cela prendrait énormément de temps pour juger toute la collection. La deuxième raison est que le jugement est subjectif. Il diffère d'un juge à un autre et aussi pour un même juge à des moments différents. Pour remédier à cela, TREC utilise la technique du pooling. Cette dernière consiste à choisir un certain nombre de runs soumis par les participants, ensuite à choisir parmi ces runs les n (généralement $n = 100$) premiers documents en éliminant les doublons. Cet ensemble de documents (appelé « pool ») est évalué par des assesseurs humains et le résultat nommé « qrels » représente la vérité terrain (perception humaine de l'utilité du documents par rapport à sa requête).

3 Approches générales pour la détection d'opinion et de la polarité

Dans la littérature, il existe deux types d'approches pour la détection d'opinion et de la polarité. Certaines sont basées sur le lexique, d'autres sur l'apprentissage.

Le premier type d'approche utilise un lexique de mots qui contiennent un sentiment. Ce lexique est soit externe c'est-à-dire construit indépendamment de tout corpus, il peut être général (SentiWordNet¹, lexique SUBJ, General Inquiry, Wilson lexicon) ou construit manuellement, soit généré automatiquement à partir du corpus (les mots qui contiennent une opinion sont extraits directement du corpus). À chaque mot du lexique est associé un ensemble de scores d'opinions et de polarité. Ce score est traité différemment par les différentes approches pour le calcul du score d'opinion d'un document. La méthode la plus simple est de donner à un document un score égal au nombre total de mots qui contiennent une opinion présents dans le document [17, 3, 18].

Le deuxième type d'approche basée sur l'apprentissage machine consiste à attribuer des données à un classifieur pour l'apprentissage. Ce dernier génère un modèle qui est utilisé pour la partie test de l'apprentissage. Ce type d'approche comprend deux aspects : extraction de features et apprentissage du classifieur. Les principales features utilisées sont : mots seuls, bigrammes, tri-grammes, part of speech (analyse de l'arbre syntaxique) et polarité. Les principaux classifieurs sont les SVM, Naive Bayes, Maximum Entropy et la régression logistique [13, 11, 8].

1. <http://sentiwordnet.isti.cnr.it/>

4 Approche proposée pour la détection d'opinions dans les blogs

Il existe des travaux qui utilisent la dépendance du topic dans leur recherche d'opinions (le titre du topic et/ou la description et/ou la narration) [11] alors que d'autres proposent de ne pas tenir compte du topic, et on parle de recherche d'opinions indépendante du topic [14, 15]. L'avantage des approches basées sur l'indépendance par rapport au topic est qu'elles sont générales et s'adaptent à n'importe quelle requête. À l'opposé, les approches basées sur les topics ont la capacité de mieux comprendre le besoin d'information de l'utilisateur vu qu'elles visent directement ce qu'il recherche. Il a été observé que les approches qui se basent sur le topic présente de meilleures performances que les approches qui sont indépendantes du topic.

Nous proposons une approche mixte basée sur le lexique SentiWordNet et l'apprentissage machine, avec un ensemble de features indépendantes des topics. Nous évaluons différentes combinaisons de ces features pour la détection d'opinions en utilisant trois classifieurs (régression logistique, SVM, Naive Bayes). Les principales features sont présentées dans ce qui suit.

1. Émotivité-1

Nous calculons l'émotivité d'un document (d) en comptant le nombre de mots (w) qui sont des adverbes (adv), adjectifs (adj) et des noms (nom) dans ce document.

$$Emot(d) = \frac{|\{w \in d | type(w) \in \{adj, adv\}\}|}{|\{w \in d | type(w) \in \{verb, nom\}\}|} \quad (1)$$

2. Émotivité-2

Cette feature est le quotient du score Émotivité-1 sur le nombre total de mots du document qui est représenté par $|\{t \in d\}|$.

$$Emot2(d) = \frac{Emot(d)}{|\{t \in d\}|} \quad (2)$$

3. Subjectivité

Cette feature utilise le lexique SentiWordNet (SWN). Dans lequel, chaque mot a plusieurs synonymes nommés « synsets ». Chaque « synset » est associé à trois scores numériques $Obj(s)$, $Pos(s)$ et $Neg(s)$, qui signifient objective, positive et négative des termes contenus dans le « synset ».

Chacun des trois scores varie de 0,0 à 1,0 et leur somme est de 1,0 pour chaque « synset ». Cela veut dire que la signification d'un « synset » est attribué au score le plus élevé des trois. Par exemple le « synset » « Estimable (3) », correspondant au sens « peut être calculé ou estimé » de l'adjectif estimable, a un score Obj de 1,0 (et les scores de Pos et de Neg est de 0,0), tandis que le synset « Estimable (1) » correspondant au sens « Digne de respect ou en haute estime » a un score positif de 0,75, un score négatif de 0,0 et un score objectif de 0,25. Dans notre méthode l'utilisation de cette ressource nous a permis d'estimer chaque mot d'un document selon l'algorithme suivant : chercher chaque mot du document dans SentiWordNet, puis calculer la subjectivité par l'équation suivante :

$$Subj(t) = \frac{\sum_{s \in t} neg(s) + pos(s)}{|t|} \quad \text{avec } t = \{s_1, \dots, s_n\} \quad (3)$$

où $|t|$ est le nombre de sens de ce terme trouvés dans le lexique SentiWordNet.

La subjectivité d'un terme t dans un document est utilisée par quatre features que nous avons proposées, et qui sont discutées ci-dessous :

– *Subjectivité-1*

$$Subj1(d) = \frac{\sum_{t \in d} Subj(t)}{|\{t \cap t' | t \in d, t' \in SWN\}|} \quad (4)$$

$|\{t \cap t' | t \in d, t' \in SWN\}|$ est le nombre total des termes trouvés dans le document qui appartiennent à SWN.

– *Subjectivité-2*

$$Subj2(d) = \sum_{t \in d} \frac{\max\left(\sum_{t \in SWN} pos(t), \sum_{t \in SWN} neg(t)\right)}{|\{t \cap t' | t \in d, t' \in SWN\}|} \quad (5)$$

Dans cette fonction, nous préférons donner un score positif à un document que si le nombre de termes positifs l'emportent dans le document et vice versa.

– *Subjectivité-3*

$$Subj3(d) = \frac{\sum_{t \in d} Subj(t)}{|\{t \in d\}|} \quad (6)$$

C'est la subjectivité-1 normalisée en la divisant par le nombre total de mots dans le document d .

– *Subjectivité-4*

$$Subj4(d) = \frac{\sum_{t \in d | Subj(t) \geq 0,5} Subj(t)}{|\{t \in d | Subj(t) \geq 0,5\}|} \quad (7)$$

La subjectivité-4 est identique à subjectivité-3, sauf que $|\{t \in d | Subj(t) \geq 0,5\}|$ est le nombre total des termes du document ayant une valeur de subjectivité supérieure à 0,5. Chacune des valeurs ci-dessus de la subjectivité est utilisée pour trois catégories de termes (adjectifs, adverbess et verbes).

4. *Réflexivité*

Les blogueurs utilisent beaucoup de pronoms réflexifs comme « je, moi, moi-même » lors de l'écriture. Par exemple, l'utilisation de « I » dans « I think », « Je pense que », « my point of view is that », « mon point de vu est que », etc. Toutes ces phrases font référence à une opinion, et par conséquent, nous incluons la mesure de la réflexivité. L'idée est que tout document avec un plus grand nombre de ces mots sera plus subjectif par rapport à celui qui en a moins. Cette mesure est exprimée par $Ref(d)$.

$$Ref(d) = \frac{|\{w \cap w' | w \in d, w' \in R\}|}{|R| + |A|} \quad (8)$$

$|\{w \cap w' | w \in d, w' \in R\}|$ est le nombre de pronoms réfléchis dans le document d qui appartiennent à la liste de réflexivité R que nous avons construit, et $|R|$ est le nombre total de pronoms dans la liste de réflexivité R , et $|A|$ est le nombre total de pronoms dans la liste d'adressage.

5. *Adressage*

La plupart des phrases trouvées dans les blogs contiennent les mots suivants « you, your, yours, yourself, yourselves, u, he, she, they, himself, herself, themselves » car les utilisateurs écrivent des commentaires sur un sujet, en s'adressant aux autres blogueurs. De ce fait l'utilisation de ces pronoms d'adressage est très fréquente. Par conséquent, nous considérons que la composante d'adressage dans le cadre de notre détection d'opinions, est comme suit :

$$Add(d) = \frac{|\{w \cap w' | w \in d, w' \in A\}|}{|A| + |R|} \quad (9)$$

$|\{w \cap w' | w \in d, w' \in A\}|$ représente le nombre d'occurrences des termes d'adressages dans le document d qui appartiennent à la liste d'adressage $|A|$ que nous avons préparée, et $|A|$ est égal au nombre total de pronoms dans la liste d'adressage A , et $|R|$ est le nombre total de pronoms dans la liste de réflexivité.

4.1 Meilleure combinaison des features pour la détection d'opinions

Nous avons réalisé la validation croisée 5-fois dans notre apprentissage machine. Cela signifie que cinq groupes de topics ont été créés A, B, C, D, E. Chaque groupe comporte 10 topics de TREC 2007. L'apprentissage machine comporte deux étapes. La première est la partie d'apprentissage, dans cette partie des données sont attribuées au classifieur pour l'apprentissage « training ». Ce dernier génère un modèle qui est utilisé pour la deuxième partie nommée « testing ».

De ce fait, un modèle est créé en utilisant le classifieur de régression logistique multinomiale [7]. Les mesures utilisées sont la précision à 10 documents (P@10) et la précision moyenne (MAP). Les résultats montrent que la combinaison de la subjectivité avec la réflexivité et l’adressage, y compris le niveau de pertinence du document (cf. tableau 1) peut être effective pour la recherche d’opinions dans les blogs. Cette combinaison a fait augmenter la mesure MAP de 4,51 %, et la P@10 de 8,61 %. Le test *t* de Student [6] montre par ailleurs que ces améliorations sont statistiquement significatives ($p < 0,05$).

Combinaison
Émotivité-1
Subjectivité-4
Réflexivité
Addressage
Score de pertinence
Rang de pertinence

TABLE 1: Combinaison des features

4.1.1 Expérimentation avec différents classifieurs

Le tableau 2 présente les résultats de la tâche de détection d’opinions en utilisant trois différents classifieurs pour la combinaison présentée dans la figure 1. Les résultats montrent que les Support Vector Machines donnent les meilleurs résultats pour cette combinaison des features alors que le modèle de régression logistique vient en deuxième position, tandis que le Naïve Bayes donne de moins bon résultats. Le résultat de 0,399 donne une amélioration de 0,66 % par rapport au résultats de la meilleure détection d’opinions qui a été faite en utilisant les topics de TREC 2007 avec une méthode qui est basée sur la dépendance des topics, et dont la MAP est égale à 0,396 [12].

Après avoir expérimenté une approche pour la détection d’opinion, nous présentons dans la section suivante nos travaux concernant l’identification de la polarité de ces opinions.

MAP	P@10	Classifieur
0,395	0,584	Logistic Regression
0,399	0,576	SVM
0,383	0,576	Naive Bayes

TABLE 2: Résultats des différents classifieurs

4.2 Détection de la polarité

La plupart des travaux liés à la détection de la polarité tiennent compte de la polarité (négative, positive ou neutre) des termes présents dans un document en utilisant différentes ressources lexicales. Un des problèmes dans ces travaux est qu’ils ne prennent pas en compte le contexte défini par le sujet recherché (la requête).

Par exemple, le mot « imprévisible » dans un document contenant des avis sur un film sera considéré comme une appréciation positive. En revanche, si le même mot est utilisé dans un autre document, par exemple pour un avis sur un appareil photo numérique, et si on dit qu'une fonctionnalité de cet appareil est imprévisible alors cette fois le mot est considéré comme un mot négatif.

Dans cette partie, nous nous intéressons à ce problème en proposant une catégorisation des topics et en étudiant son impact sur la détection de la polarité. L'approche de la détection de polarité proposée est une approche mixte utilisant le lexique SentiWordNet et l'apprentissage machine basé sur le classifieur de type Régression Logistique.

4.2.1 Catégorisation des topics

Nous avons classé les sujets de TREC Blog 2006 et TREC Blog 2007 dans les 6 classes mentionnées ci-dessous :

1. Film / TV ;
2. Personne ;
3. Organisation ;
4. Événement ;
5. Produit ;
6. Thème.

Cette classification est inspirée de celle des travaux de [17]. Chaque thème de TREC Blog 2006 et TREC Blog 2007 a été lu par deux personnes que nous avons recrutées (appelés annotateurs). Le tableau 3 montre l'accord des résultats entre les deux annotateurs. Dans les instructions aux annotateurs, nous leur avons demandé de lire attentivement les descriptions des sujets, puis le titre de chaque sujet et de leur attribuer une classe parmi les classes proposées. L'étude de ce tableau a montré que l'on trouve très peu de désaccords entre les annotateurs, la plupart des désaccords se trouvent dans les sujets de l'année 2007 (15 désaccords par rapport à seulement 1 pour l'année 2006).

Annotateur 1	Annotateur 2						
	FILM	PERS	ORG	EVENT	PROD	THEM	TOTAL
FILM	12	0	0	3	0	1	16
PERS	0	20	1	0	0	0	21
ORG	0	0	14	3	0	2	19
EVEN	0	1	0	7	0	2	10
PROD	0	0	1	0	13	3	17
THEME	0	0	0	2	0	15	17
TOTAL	12	21	16	15	13	23	100

TABLE 3: Jugement des annotateurs pour les différents thèmes

La mesure statistique κ de Cohen [2] a été utilisée pour calculer le score d'accord entre les annotateurs. La mesure *Kappa* est connue pour être une mesure plus robuste que le calcul de pourcentage, car elle tient compte de l'accord qui se produit par hasard. Son équation est la suivante :

$$\kappa = \frac{Pr(a) - Pr(e)}{1 - Pr(e)} \quad (10)$$

où $Pr(a)$ est la proportion observée d'accord entre annotateurs, et $Pr(e)$ est la probabilité d'un accord hypothétique par chance. La valeur κ peut varier de -1 (totale contradiction) à 1 (plein accord). Selon notre travail la valeur κ a été calculée par l'équation (10), le résultat de $0,77$ représente un accord entre les deux annotateurs. Concernant les autres topics pour lesquels les annotateurs n'ont pas été en accord, un troisième annotateur a résolu le conflit.

Le tableau 4 ainsi que le tableau 5 montrent les différentes classes attribuées à chaque topic de 2006 et 2007.

Classe	Topics 2006	TOTAL
EVEN	853, 861, 890	3
THEME	855, 858, 859, 865, 867, 868, 869, 878, 889, 894, 896, 898, 895	13
FILM	851, 860, 864, 872, 875, 876, 881, 886, 895	9
PERS	852, 854, 857, 870, 871, 873, 874, 880, 891, 892, 897	11
PROD	856, 862, 879, 883, 893	5
ORG	863, 866, 877, 882, 884, 885, 887, 888, 900	9

TABLE 4: Classification en thèmes des topics de TREC Blog 2006

Classe	Topics 2007	TOTAL
EVEN	943, 938, 936, 933, 925, 923, 914, 913, 906, 905	10
THEME	901, 902, 907, 918, 921, 927, 929, 931, 942	9
FILM	911, 922, 928	3
PERS	903, 904, 908, 920, 924, 935, 940, 941, 947, 949	10
PROD	909, 916, 917, 932, 934, 937, 939, 944, 946, 950	10
ORG	910, 912, 915, 919, 926, 930, 945, 948	8

TABLE 5: Classification en thèmes des topics de TREC Blog 2007

Dans ce qui suit, nous effectuons des expériences sur la détection de polarité en utilisant cette catégorisation des topics. Nous allons ensuite analyser les effets de cette catégorisation. Pour cela nous effectuons des expériences en deux phases. Dans la première phase, nous réalisons des expériences sur la détection de polarité sans classification des topics et dans la seconde phase, nous re-expérimentons avec des topics de TREC 2006 et 2007 en utilisant la classification des topics.

4.3 Détection de polarité sans classification des topics

Nous utilisons un modèle de régression logistique pour nos expériences. Les features utilisées pour détecter la polarité sont : le nombre de mots positifs, le nombre de mots négatifs, le nombre de mots neutres, et le nombre d'adjectifs dans un document. Nous avons, dans un premier temps, choisi des features ordinaires et simples (déjà utilisées dans [5]). Il est à rappeler que ce travail ne vise pas à proposer de nouvelles fonctionnalités permettant de détecter la polarité des documents, mais de vérifier l'impact de la classification des topics dans la polarité. Pour déterminer la polarité, il nous a fallu une base de départ avec des documents qui contiennent des opinions. Pour cela nous utilisons le run que nous avons obtenu par la feature subjectivité-4 (dont la MAP est de 0,387 et la précision P@10 de 0,546) comme point de départ. Les expérimentations pour la détection de polarité sans classification ont été effectuées dans trois environnements différents. Toutes les expériences avec leurs paramètres sont expliquées ci-dessous.

1. *expérimentation 1*

Les expériences ont été effectuées en utilisant les mêmes features que celles expliquées ci-dessus. Une validation croisée à 5 a été réalisée pour les sujets de l'année 2007. Le résultat pour la mesure MAP Positive est de 0,099 et pour la mesure de précision P@10 de 0,200. Pour la mesure MAP Négative, le résultat est égal à 0,065 et pour P@10, il est égal à 0,060.

Problème avec cette expérimentation

Dans cette expérimentation le volume de données utilisées dans la partie d'apprentissage est beaucoup plus grand que celui des données utilisées pour la classification. Par conséquent, avant de discuter d'autres causes qui pourraient améliorer ces résultats, nous procédons à une autre expérience en utilisant un nombre réduit de données d'entraînement pour des expériences sans classification de topics.

2. *expérimentation 2*

Dans ce contexte, les données d'apprentissage ont été ramenées de 40 à 22 topics. 22 est le nombre maximum de sujets d'un groupe dans l'expérience avec classification de topics, par conséquent, nous avons réduit le nombre de sujets à 22 pour l'apprentissage. Le résultat obtenu pour la MAP Positive est de 0,163 et pour la précision p@10 de 0,200. La MAP Négative est égale à 0,062 et la P@10 est de 0,058.

Problème avec cette expérimentation

Le choix des sujets d'essai pour cette expérience a été fait de manière aléatoire. Le choix des topics d'essai a été fait dans l'ordre numérique : le premier essai a été fait pour les sujets de 901 à 910, le deuxième pour les sujets de 911 à 920 tandis que les topics d'apprentissage ont été choisis au hasard. Le problème qui peut donc se poser est que des topics de la même classe peuvent être dans la partie test et dans la partie apprentissage, ce qui devrait être évité. Par conséquent, nous effectuons une autre expérimentation en utilisant un troisième paramètre.

3. *expérimentation 3*

Pour résoudre le problème de la deuxième expérimentation et améliorer davantage la détection de la polarité, nous avons différencié les topics pour le test et les topics pour l'apprentissage. Pour les données de test, nous utilisons l'ensemble des sujets de la classe « Event » (ou de la classe « produit » ou d'une autre classe) et pour les données d'apprentissage, nous utilisons 22 autres sujets qui n'appartiennent pas à la classe relative aux données d'essai. Ce paramètre a également donné de bons résultats pour l'expérimentation avec classification des topics mais une grosse chute dans la MAP a été constatée. Il est donc important de souligner que la présence des topics de la même classe dans la partie apprentissage et dans la partie test a un impact considérable sur les performances. Le résultat est pour la MAP

Positive MAP de 0,055, et pour la P@10 de 0,072 tandis que la MAP Négative est égale à 0,036 et la mesure de P@10 est égale à 0,054. Le tableau 6 récapitule les résultats des trois expérimentations.

RUN	POS		NEG	
	MAP	P@10	MAP	P@10
Expérimentation 1	0,099	0,200	0,065	0,060
Expérimentation 2	0,163	0,200	0,062	0,058
Expérimentation 3	0,055	0,072	0,036	0,054

TABLE 6: Résultats des différentes expérimentations de détection sur la polarité

4.4 Détection de la polarité avec classification des topics

Dans cette partie nous utilisons les mêmes features que celles utilisées pour la détection de la polarité sans classification : nombre de mots positifs, nombre de mots négatifs, nombre de mots neutres, ainsi que le nombre d'adjectifs. Un groupe de topics a été créé pour chaque classe. Nous avons considéré les topics de TREC 2006 et TREC 2007. Une validation croisée à n a été réalisée parmi les sujets de chaque groupe, en utilisant un modèle de régression logistique, où n est le nombre de sujets dans un groupe. Nous nous sommes reportés uniquement aux résultats des topics de TREC 2007.

Nous avons inclus les sujets de TREC 2006, afin d'augmenter les données d'entraînement pour l'apprentissage machine de sorte que la comparaison entre « avec classification des topics » et « sans classification des topics » peut être justifiée. Les résultats de cette expérimentation sont donnés dans le tableau 7.

	POS		NEG	
	MAP	P@10	MAP	P@10
Sans classification	0,055	0,072	0,036	0,054
Avec classification	0,109	0,146	0,068	0,068
% amélioration	98,18	102,77	85,24	26,87

TABLE 7: Comparaison des résultats de la polarité

Ce tableau présente les résultats de la détection de polarité selon la classification des topics pour les polarités positives et négatives. Ces résultats montrent que la classification des topics a bien amélioré les résultats pour toutes les expérimentations réalisées. Nous ne commenterons que la dernière expérience, qui est considérée comme la plus adaptée à notre travail. Une amélioration considérable peut être notée dans les résultats, cela est dû au choix de notre baseline qui est très faible. Toutefois, ces résultats sont suffisants pour prouver que la catégorisation des topics peut améliorer les résultats de la détection de la polarité. Il est à noter que l'objectif de ce travail n'était pas d'améliorer le travail des autres travaux sur la détection de la polarité, mais plutôt d'analyser les effets de la classification sur la tâche de détection d'opinions.

5 Conclusion

Dans une première étape, nous nous sommes intéressés à la détection d'opinions dans les blogs. Nous proposons pour cela une approche mixte basée sur le lexique et l'apprentissage machine. Les features que nous avons utilisées sont indépendantes des topics. Une étude a été effectuée pour trouver la combinaison des features améliorent le plus la détection d'opinion à l'aide du classifieur Régression Logistique. Cette combinaison des features a été testée sur les classifieurs SVM et Naive Bayes, les résultats ont montré que le SVM est celui qui améliore au mieux la détection d'opinions. En utilisant la méthode mixte proposée, la combinaison des features trouvée (Emotivité-1, Subjectivité-4, Reflexivité, Addressage, score de pertinence et le rang de pertinence) et le classifieur SVM, les résultats obtenus améliorent de 5,45 % la baseline-4 de TREC 2008 (la plus élevée et donc la plus difficile à améliorer).

Dans une deuxième étape, nous proposons une approche prenant en considération le contexte pour la détection de la polarité. L'approche est mixte : lexique et apprentissage machine du type régression logistique. Une étude de l'impact de la catégorisation des topics de TREC Blog 2007 a été effectuée. Il en résulte que la polarité est très sensible à la catégorisation : une amélioration de 98,18 % pour la polarité positive et de 85,24 % pour la polarité négative a été constatée et ceci comparativement aux performances obtenues sans classification.

Références

- [1] C. Buckley and E. M. Voorhees. Retrieval System Evaluation. In E. M. Voorhees and D. K. Harman, editors, *TREC : Experiment and Evaluation in Information Retrieval*, chapter 3, pages 53–75. MIT Press, Cambridge, MA, USA, 2005.
- [2] J. Cohen. A Coefficient of Agreement for Nominal Scales. *Educ. Psychol. Meas.*, 20 :37–46, 1960.
- [3] C. Fautsch and J. Savoy. UniNE at TREC 2008 : Fact and Opinion Retrieval in the Blogosphere. In *TREC : Proceedings of the Text Retrieval Conference*, 2008.
- [4] D. Hannah, C. Macdonald, J. Peng, B. He, and I. Ounis. University of Glasgow at TREC 2007 : Experiments in Blog and Enterprise Tracks with Terrier. In *TREC : Proceedings of the Text Retrieval Conference*, 2007.
- [5] L. Hoang, S.-W. Lee, G. Hong, J.-Y. Lee, and H.-C. Rim, editors. *A Hybrid Method for Opinion Finding Task (KUNLP at TREC 2008 Blog Track)*, 2008.
- [6] D. Hull. Using Statistical Testing in the Evaluation of Retrieval Experiments. In *SIGIR'93 : Proceedings of the 16th annual international ACM SIGIR conference*, pages 329–338, New York, NY, USA, 1993. ACM Press.
- [7] D. T. Larose, editor. *Data Mining Methods and Models*. Wiley, 2006.
- [8] Y. Lee, S.-H. Na, J. Kim, S.-H. Nam, H.-Y. Jung, and J.-H. Lee. KLE at TREC 2008 Blog Track : Blog Post and Feed Retrieval. In *TREC*, 2008.
- [9] C. Macdonald, I. Ounis, and I. Soboroff, editors. *Overview of the TREC 2007 Blog Track*, 2008.
- [10] C. Macdonald, I. Ounis, and I. Soboroff, editors. *Overview of the TREC 2009 Blog Track*, 2010.
- [11] G. Mishne and M. de Rijke. A study of blog search. In *ECIR*, pages 289–301, 2006.
- [12] R. L. T. Santos, B. He, C. Macdonald, and I. Ounis. Integrating proximity to subjective sentences for blog opinion retrieval. In *ECIR*, pages 325–336, 2009.
- [13] R. Song, Q. Tang, D. Shi, H. Lin, and Z. Yang. DUTIR at TREC 2007 Blog Track. In *TREC : Proceedings of the Text Retrieval Conference*, 2007.

- [14] E. M. Voorhees and L. P. Buckland, editors. *Proceedings of The Sixteenth Text REtrieval Conference, TREC 2007, Gaithersburg, Maryland, USA, November 5-9, 2007*, volume Special Publication 500-274. National Institute of Standards and Technology (NIST), 2007.
- [15] E. M. Voorhees and L. P. Buckland, editors. *Proceedings of The Seventeenth Text REtrieval Conference, TREC 2008, Gaithersburg, Maryland, USA, November 18-21, 2008*, volume Special Publication 500-277. National Institute of Standards and Technology (NIST), 2008.
- [16] K. Yang, N. Yu, and H. Zhang. WIDIT in TREC 2007 Blog Track : Combining Lexicon-Based Methods to Detect Opinionated Blogs. In *TREC : Proceedings of the Text Retrival Conference, 2007*.
- [17] G. Zhou, H. Joshi, and C. Bayrak. Topic categorization for relevancy and opinion detection. In *TREC : Proceedings of the Text Retrieval Conference, 2007*.
- [18] O. Zubaryeva and J. Savoy. Opinion and Polarity Detection within Far-East Languages in NTCIR-7. 2008.