
Un modèle de recherche d'information sociale dans les microblogs : cas de Twitter

Lamjed Ben Jabeur, Lynda Tamine et Mohand Boughanem

IRIT, Université Paul Sabatier

118 Route de Narbonne, F-31062 TOULOUSE CEDEX 9

jabeur, tamine, boughanem@irit.fr

RÉSUMÉ. Cet article propose une approche sociale pour la recherche d'information dans les microblogs. Nous étudions en particulier le réseau social de Twitter et nous modélisons ses principales entités et relations. En outre, nous proposons de représenter l'influence au sein du réseau à travers les relations de rediffusion des messages et nous introduisons un modèle de recherche d'information sociale qui évalue la pertinence en tenant compte de l'influence du blogueur et de son expertise par rapport au sujet de la requête. Plus précisément, notre modèle combine la pertinence thématique des publications avec un premier score d'influence sociale obtenu par l'application de l'algorithme PageRank sur le réseau social de rediffusion et un deuxième score d'expertise obtenu par un modèle de langue. Nous évaluons l'efficacité de notre modèle sur une collection de 237000 microblogs publiés sur Twitter. Les résultats obtenus montrent l'intérêt de l'intégration de la structure du réseau dans l'évaluation de l'importance des blogueurs.

ABSTRACT. We present in this paper a social approach for information retrieval over microblogs. We study in this work the social network of Twitter and we model the main involved entities and relationships. In addition, we model the influence within the network through the retweeting behavior and we introduce a social information retrieval model that evaluates relevance by considering both the influence and the expertise of corresponding bloggers regarding the query topic. Specifically, our model combines the topical relevance of tweets with a first score of influence computed by applying the PageRank algorithm on the retweet network and a second score of expertise computed based on a language model. We evaluate the effectiveness of our model on a microblogs dataset including 237000 microblogs published in Twitter. Obtained results show the interest in integrating the social network structure for evaluating the social importance of bloggers.

MOTS-CLÉS : Recherche d'information sociale, réseaux sociaux, Twitter, rediffusion, influence, expertise, importance sociale.

KEYWORDS: Social information retrieval, social network, Twitter, retweet, influence, expertise, social importance.

1. Introduction

Un service de *microblogage* est à la fois un moyen de communication et un système de collaboration qui permet le partage et la diffusion des messages textuels. Il permet aux utilisateurs de communiquer des informations sur leurs statuts, activités, pensées et opinions [JAV 09]. Au-delà de ces aspects d'utilisation personnelle à des fins de divertissement, les *microblogs* offrent aux entreprises et aux communautés virtuelles un moyen de collaboration rapide et pratique. En outre, les *microblogs*, vus comme une nouvelle source d'information, commencent à concurrencer les médias de masse. Ils donnent à tout utilisateur la possibilité de rapporter une actualité qu'il a vécue ou dont il a été témoin selon son point de vue et ce, d'une manière plus démocratique.

En comparaison aux services de *microblogage* disponibles sur le Web, Twitter¹ reste le site le plus populaire avec plus de 190 millions de visiteurs par mois². Vu ce succès, la quantité de données issues par les *microblogs* a considérablement augmenté avec un taux qui excède les 60 millions de publications par jour. Ainsi, les utilisateurs trouvent une difficulté pour accéder aux dernières actualités, masquées par l'énorme quantité des données et le flux soutenu des publications.

La recherche d'information dans les *microblogs* est particulièrement limitée par la taille courte des articles qui augmente à son tour la difficulté de la recherche textuelle par mots-clés. Les travaux dans ce domaine s'orientent alors vers l'intégration d'autres contextes tels que le cadre spatio-temporel [SAN 09] [SAK 10] et la structure du réseau social [WEN 10]. Dans cette dernière catégorie d'approches, la recherche d'information sociale permet d'exploiter les interactions sociales et la topologie du réseau afin d'évaluer l'utilité des articles et cela à travers l'importance des acteurs associés.

Dans cet article, nous proposons d'exploiter la structure du réseau social pour la recherche d'information dans les *microblogs* et nous introduisons un modèle social qui combine la pertinence thématique de l'article et l'importance sociale du *blogueur*. Ce modèle représente le réseau social des *blogueurs* au moyen d'un réseau d'influence et associe l'importance sociale à la capacité d'influence ainsi que l'expertise du sujet de la requête. L'influence des *blogueurs* est évaluée dans ce modèle par l'application de l'algorithme *PageRank* sur le réseau de rediffusion des articles.

Cet article est organisé comme suit. Dans la section 2, nous décrivons le réseau de *microblogage* de Twitter. Dans la section 3, nous présentons la tâche de recherche de *tweets* et nous discutons certains travaux proposés dans ce domaine. Dans la section 4, nous modélisons le réseau social d'influence et nous introduisons notre modèle de recherche d'information dans les *microblogs* basé sur l'influence et l'expertise des *blogueurs*. Dans la section 5, nous menons une série d'expérimentations pour valider notre approche. Enfin, la section 6 conclut l'article et annonce des perspectives.

1. <http://twitter.com/>

2. <http://techcrunch.com/2010/06/08/twitter-190-million-users/>

2. Le réseau social d'information de Twitter

À la différence des autres services de *microblogage*, Twitter se positionne par la relation sociale d'*abonnement*. Cette association dirigée, permet aux utilisateurs d'exprimer leur intérêt pour les articles d'un autre microblogueur.

Le réseau social de Twitter ne se limite pas aux *blogueurs* et aux relations d'abonnement, il inclut également tous les acteurs et les données qui interagissent entre eux dans les deux contextes de publication et d'utilisation des articles. Nous avons analysé le service de *microblogage* de Twitter et nous avons dégagé, comme illustré dans la figure 1, les différentes entités impliquées dans le réseau d'information et les diverses relations qui les associent.

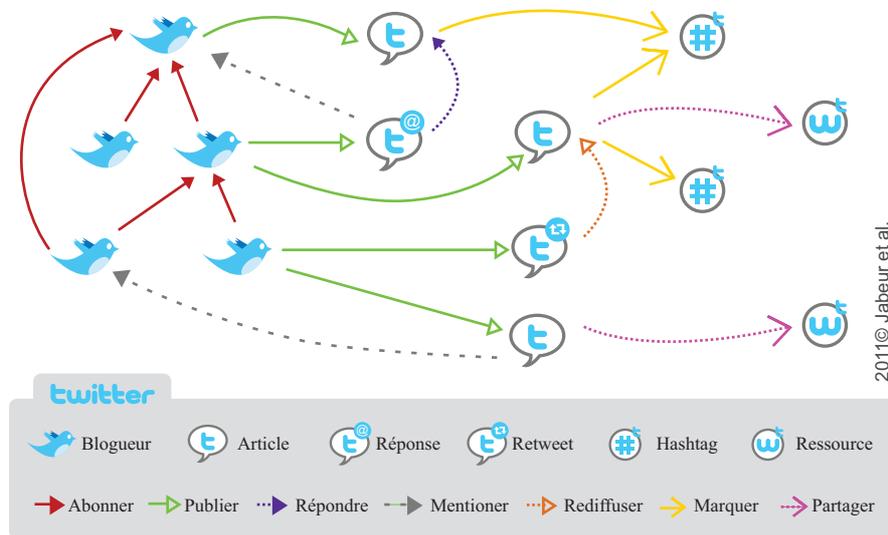


Figure 1 – Réseau social d'information de Twitter

3. La recherche d'information dans le réseau de Twitter

La recherche des *tweets* est une tâche de recherche d'information *ad-hoc* dont l'objectif est de sélectionner les articles des *microblogs* pertinents en réponse à une requête Q [OUN 10]. La définition de la pertinence dans la recherche des *tweets* ne se limite pas à la similarité textuelle mais elle tient compte également des interactions sociales dans le réseau. Dans ce contexte, la pertinence des articles dépend aussi de l'importance du *blogueur* qui les soumet.

En comparaison à la recherche Web, la recherche des *tweets* permet d'obtenir une information brève, concise et précise sur un sujet actuel [JAV 09]. Elle permet aussi de recevoir en temps réel des informations sur un événement qui vient de se produire

quelques secondes avant [TEE 11]. Enfin, la recherche des *tweets* permet d'accéder aux actualités avec une diversité de points de vue des *blogueurs* et à une échéance proche de l'événement [SAK 10].

La recherche des *tweets* est une tâche difficile et présente plusieurs problématiques à résoudre. Parmi ces problèmes, nous identifions l'indexation en temps réel de flux soutenus des articles [SAN 09] ainsi que le problème des *spams* et des messages parasites [SAK 10]. Dans ce travail, nous adressons principalement les problèmes de la mesure d'importance du *blogueur* et de l'intégration des facteurs de pertinence.

En ce qui concerne l'importance de blogueurs, de nombreux travaux ont proposé d'étudier la qualité des blogueurs comme une première étape pour estimer la pertinence de leurs articles. Nous notons qu'un article publié par un *blogueur* important est plus pertinent dans ce contexte qu'un autre article écrit par un "inconnu". Dans l'objectif de mesurer l'importance d'un *blogueur*, [BAL 08] propose d'évaluer son expertise au sujet de la requête et cela en se basant sur un modèle de langue. D'autres approches telles que [ZHA 07] et [NOL 09] considèrent que les experts d'un domaine sont connectés par des relations sociales et proposent d'explorer la topologie du réseau social afin de les identifier.

Au delà de l'expertise, les travaux de [KWA 10] et [AGA 08] proposent de classer les *blogueurs* selon leur pouvoir d'influence. Il est constaté que la rediffusion des messages est souvent corrélée à l'influence des *blogueurs* et non pas au nombre des abonnés, la popularité ou l'autorité de *blogueurs*. Dans ce contexte, *TwitterRank* [WEN 10] propose de mesurer l'influence des *blogueurs* par application de l'algorithme *Page-Rank* sensible à la thématique sur le réseau d'abonnement.

Pour l'intégration des facteurs de pertinence, certaines approches modulaires proposent de calculer de multiples facteurs de pertinence en se basant sur la structure du réseau et de les combiner ultérieurement [MUT 01] [KIR 06]. Ces approches étudient l'importance sociale en appliquant les mesures d'analyse des réseaux sociaux et estiment une pertinence globale par la combinaison de la pertinence thématique et de l'importance sociale. D'autres approches intégrées modélisent tous les facteurs de pertinence sur le réseau et estiment la pertinence globale par une probabilité de transition [AME 07] [ZAI 10].

Nous proposons dans cet article un modèle pour la recherche sociale des *tweets* qui associe la pertinence des articles à l'importance sociale des *blogueurs* correspondants. Les deux problèmes majeurs qui s'imposent dans cette approche sont la modélisation du réseau social et l'identification des *blogueurs* importants et des sources d'informations pertinentes. Comparativement aux autres approches dans le domaine nous proposons :

- d'estimer l'importance d'un *blogueur* à travers son influence et son expertise sur le sujet de la requête. Les travaux similaires évaluent l'importance d'un *blogueur* uniquement par son expertise [BAL 08] ou bien par son influence sur le réseau [KWA 10] [AGA 08].

- de modéliser le réseau social des *blogueurs* en se basant sur les relations de rediffusion à la différence des approches proposées dans [WEN 10] [KWA 10] qui se limitent aux relations d’abonnement pour la représentation du réseau social de Twitter.
- de prendre en compte la pertinence sociale d’expertise et d’influence et les combiner avec la pertinence thématique afin d’évaluer une pertinence globale. Certains travaux sur les *microblogs* se basent sur la pertinence thématique et les facteurs spatio-temporels sans considérer le contexte social [SAN 09] [SAK 10].

4. Un modèle social basé sur l’influence des blogueurs pour la recherche d’information dans les microblogs

Nous introduisons ici un modèle de recherche des *tweets* qui combine la pertinence thématique et l’importance sociale des *blogueurs*. Ce modèle considère l’influence et l’expertise comme les principaux facteurs sociaux qui déterminent l’importance du *blogueur* et la qualité de ses articles. Nous notons que l’influence d’un *blogueur* dépend de ses relations de rediffusion et elle est estimée selon sa position dans le réseau social d’influence. D’autre part, l’expertise est déterminée par la distribution des termes dans ses articles suivant un modèle de langue. Nous décrivons dans la suite la structure du réseau social d’influence sur laquelle nous nous basons pour le calcul du score d’influence et nous présentons notre approche pour la combinaison des différents facteurs de pertinence.

4.1. Le réseau social d’influence

Le *blogueur* confirme par la retransmission d’un article, l’importance du message communiqué. Il montre également qu’il s’intéresse à son sujet et qu’il adopte la même idée si une opinion y est exprimée. D’autre part, il est constaté que les abonnés continuent à rediffuser les messages s’ils jugent leur contenu est important. L’importance d’un *blogueur* est alors déterminée par la proportion de ses messages rediffusés. Cela exprime aussi son pouvoir d’influence sur le réseau social.

Afin d’évaluer l’influence, nous proposons de modéliser le réseau social de Twitter en se basant sur les relations de rediffusion. Le réseau social d’influence est modélisé par un graphe $G = (U, E)$ où U est l’ensemble des utilisateurs et $E = U \times U$ représente l’ensemble des relations d’influence entre eux. Une relation d’influence $e(u_i, u_j) \in E$ est définie de $u_i \in U$ vers $u_j \in U$ si et seulement s’il existe au moins un article publié par u_j et rediffusé par u_i . Nous calculons le poids $w(u_i, u_j)$ de la relation d’influence par la formule suivante :

$$w(u_i, u_j) = \frac{\text{nb articles publiés par } u_j \text{ et rediffusés par } u_i}{\text{nb articles rediffusés par } u_i} \quad (1)$$

Nous présentons dans la figure 2 un exemple de réseau d’information sur Twitter à partir duquel nous extrayons le réseau social d’influence. À gauche de la figure, nous

modélisons le réseau social d'information. Le réseau social d'influence est présenté à la droite de la figure en utilisant la notation des graphes. Les nœuds correspondent aux utilisateurs et les arcs représentent les relations d'influence.

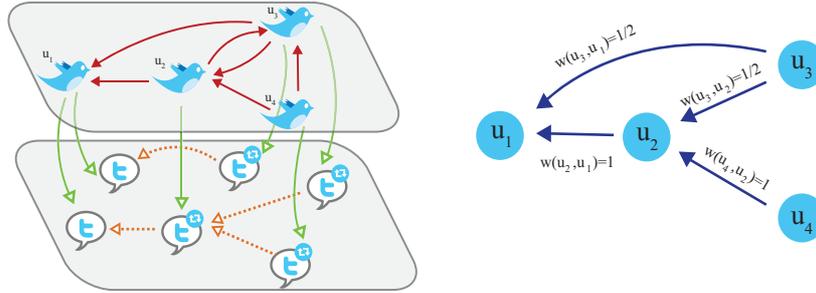


Figure 2 – Extraction du réseau social d'influence

4.2. Vers l'estimation de la pertinence sociale

La recherche des *tweets* est une tâche de recherche d'information *ad-hoc* dont l'objectif est de sélectionner les articles pertinents en réponse à une requête utilisateur. Afin de présenter une liste précise d'articles, notre modèle combine un score de pertinence thématique et un score de pertinence sociale. L'objectif de cette combinaison est de présenter une liste d'articles qui couvrent le sujet de la requête et qui sont publiés par des *blogueurs* importants. Ces deux scores sont combinés linéairement selon la formule suivante :

$$Rel(Q, t, G) = \alpha RSV(Q, t) + (1 - \alpha) S(Q, u_t, G) \quad (2)$$

Où Q , t et G représentent respectivement la requête, l'article et le réseau d'influence social. $\alpha \in [0..1]$ est un paramètre de pondération. $RSV(Q, t)$ est le score normalisé de la pertinence thématique. $S(Q, u_t, G)$ est le score normalisé de l'importance sociale avec u_t qui correspond au *blogueur* ayant publié l'article t .

Nous notons que la pertinence thématique $RSV(Q, t)$ dépend uniquement de l'article et de la requête et nous proposons d'utiliser le modèle probabiliste Okapi Bm25 [ROB 98] pour le calcul de ce score. Concernant l'importance sociale $S(Q, u_t, G)$, nous précisons que ce score prend en compte la structure du réseau G . Dans la suite, nous détaillons le calcul de ce score.

4.3. L'importance sociale des blogueurs

Le score d'importance sociale $S(Q, u, G)$ évalue l'importance d'un *blogueur* u en se basant sur sa position dans le réseau social G . En plus de son influence sociale,

nous proposons d'évaluer l'expertise du *blogueur* sur le sujet de la requête. Par conséquent, seuls les auteurs influents et expérimentés seront pris en considération à l'exception des auteurs caractérisés d'une influence absolue, elle-même indépendante de la requête. Grâce à la prise en compte de l'expertise de l'auteur, l'évaluation de l'importance sociale se limite au sous-réseau des *blogueurs* relatif au sujet de la requête. Cela évite le problème de dérivation du sujet et la domination des certains *blogueurs*.

L'importance sociale des *blogueurs* est calculée par la multiplication de deux scores d'influence et d'expertise comme le décrit la formule suivante :

$$S(Q, u, G) = Inf_G(u) * Exp(Q, u) \quad (3)$$

Où $Inf_G(u)$ mesure l'influence du *blogueur* u dans le réseau social G et le score $Exp(q, u)$ mesure son expertise au sujet de la requête Q .

4.3.1. L'influence des blogueurs

Nous calculons le score d'influence par l'application de l'algorithme *PageRank* sur le réseau social d'influence. Par analogie avec le principe d'autorité des pages Web, l'influence d'un *blogueur* provient du vote des autres *blogueurs* avec lesquels il partage une relation d'influence. Plus ces *blogueurs* sont importants et reçoivent à leur tour des liens d'influence, plus le *blogueur* est considéré influent. Le score d'influence d'un *blogueur* est calculé sur plusieurs itérations avec la formule suivante :

$$Inf_G(u_i) = d \frac{1}{|U|} + (1 - d) \sum_{u_j: e(u_j, u_i) \in E} w(u_j, u_i) \frac{Inf_G(u_j)}{O(u_j)} \quad (4)$$

Avec $O(u_j)$ est le nombre de relations d'influence à partir de l'utilisateur u_j , $d \in [0, 1]$ est un paramètre de configuration de l'algorithme *PageRank* et $w(u_j, u_i)$ est le poids de la relation d'influence entre u_j et u_i comme défini dans l'équation 1. Ce poids permet d'ajuster la probabilité de transition par défaut et de favoriser les fortes influences dans le réseau.

4.3.2. L'expertise des blogueurs

Afin d'évaluer son expertise, nous représentons chaque *blogueur* par un profil \mathcal{B}_u . Ce profil correspond à l'ensemble de son microblog et inclut tous ses messages. Les articles de chaque *blogueur* sont représentés par un super document dont l'ensemble constitue la collection des profils des *blogueurs*. Nous appliquons un modèle de langue sur cette collection et nous trions les supers documents des profils selon leur pertinence à la requête. Le score obtenu de chaque profil correspond à l'expertise du *blogueur*. Ce score est calculé selon la formule suivante :

$$Exp(q, u) = \prod_{t \in q} ((1 - \lambda)p(t|\mathcal{B}_u) + \lambda p(t))^{n(t, q)} \quad (5)$$

Où les probabilités $p(t|\mathcal{B}_u)$ et $p(t)$ correspondent respectivement à la probabilité d'apparition d'un terme t sachant le super document \mathcal{B}_u et la probabilité d'apparition d'un

terme t dans la collection. Ces probabilités sont calculées par un maximum de vraisemblance comme défini dans [Hie 01]. Enfin, $n(t, q)$ représente le nombre d'occurrences du terme t dans la requête q .

5. Evaluation expérimentale préliminaire

Afin de valider notre modèle pour la recherche des *tweets*, nous avons effectué une série d'expérimentations préliminaires sur une collection d'articles de Twitter. Les deux objectifs principaux de cette évaluation sont :

- Comparer les différentes composantes de la pertinence thématique, de l'influence sociale et de l'expertise.
- Etudier l'impact de la combinaison du score thématique et du score social sur l'estimation globale de la pertinence.

5.1. Description de la collection de test

Avec l'absence d'un cadre standard pour l'évaluation de la recherche d'information dans les *microblogs*, nous avons construit une collection d'articles et de requêtes. Nous décrivons dans la suite la collection d'articles, le réseau social et la démarche adoptée pour collecter les jugements de pertinence.

- Collection d'articles : nous avons construit une collection d'articles de Twitter qui inclut outre le contenu des messages, des métadonnées sur les relations d'abonnement et de rediffusion. Cette collection est collectée à l'aide de l'API Twitter³ qui permet d'accéder aux flux d'actualités et aux informations sociales de chaque *blogueur*. Cette collection inclut au total 237000 articles Twitter publiés par 2400 *blogueurs* qui s'intéressent aux technologies d'Internet. Elle est construite selon 3 étapes :

1) En se basant sur l'annuaire WeFollow⁴ qui classe les *blogueurs* par centres d'intérêt, nous avons sélectionné un ensemble des *blogueurs* dans la catégorie des technologies d'internet. Nous vérifions s'il existe un compte Twitter avec le même identifiant et qu'il est encore actif.

2) À l'aide d'API Twitter, nous avons collecté les 100 derniers articles de chaque *blogueur* ainsi que son réseau d'abonnement. Les données collectées incluent aussi les références vers les articles d'origine dans le cas où le message est rediffusé.

3) À partir de la liste des *blogueurs* sélectionnés, nous gardons seulement ceux qui ont un accès public à leurs articles et nous prenons en compte seulement les relations sociales dont les deux *blogueurs* font partie de la collection des articles.

3. <http://dev.twitter.com/>

4. <http://wefollow.com/>

- **Réseau social** : nous présentons dans le tableau 1 des statistiques sur le réseau social extrait de la collection d'articles. Nous notons qu'une grande majorité des *blogueurs* ont un nombre équitable d'articles qui dépasse 98 en moyenne. Nous constatons aussi que le réseau social d'abonnement est plus dense avec un degré moyen de 31.89 comparé au réseau d'influence d'un degré moyen de 1.23. Cependant, le réseau d'influence présente une composante géante qui couvre 79.5% du réseau comme le montre la figure 3. Cette composante géante couvre la majorité du réseau et montre donc la connectivité et l'homogénéité du réseau d'influence.

Blogueurs	2400
Articles	237000
→ Nb moyen d'articles	98.75
Relations d'abonnement	76 542
→ Degré moyen	31.89
Relations d'influence	1231
→ Degré moyen	1.31

Tableau. 1 – Statistique sur la collection

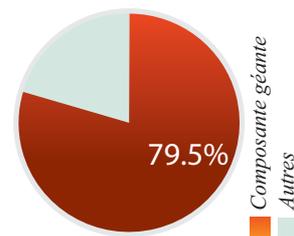


Figure. 3 – Composante la plus géante du réseau

- **Requêtes et jugements de pertinence** : afin de construire la collection des requêtes et des jugements de pertinence, nous avons demandé à 5 utilisateurs de définir chacun 2 requêtes sur des actualités récentes dans le domaine des technologies d'Internet et de juger la pertinence de 100 articles pour chaque requête. Les 100 articles dont a été jugée la pertinence sont sélectionnés sur la base de leur pertinence thématique. Pour cela, nous avons indexé la collection avec le moteur de recherche Terrier⁵ et pour chaque requête nous avons trié les articles en utilisant la formule Okapi BM25 comme schéma de pondération. Nous avons collecté au total 106 articles pertinents sur un ensemble de 10 requêtes.

5.2. Comparaison des facteurs de pertinence

Nous évaluons dans cette expérimentation les facteurs de pertinence sociale à savoir l'expertise et l'influence et nous comparons leurs performances de façon isolée. La figure 4 montre les valeurs de MAP (Mean Average Precision) obtenues par les différentes configurations de notre modèle. Nous précisons que ces 3 configurations ne tiennent pas compte de la pertinence thématique et elles sont obtenues avec $\alpha = 0$ dans la formule 2. Dans la première configuration notée *Expertise*, seule l'expertise est prise en compte dans le calcul de l'importance sociale de *blogueur* $S(q, u, G) = Exp(S, u)$. La seconde configuration notée *Influence* prend en compte

5. <http://terrier.org/>

seulement le score d'influence $S(q, u, G) = Inf_e(u)$. Finalement, le modèle *Expertise & Influence* correspond au score d'importance sociale du *blogueur* comme défini dans la formule 3.

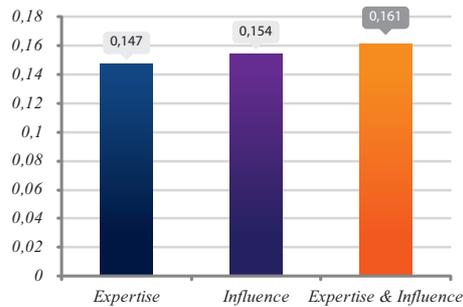


Figure 4 – Comparaison des facteurs de pertinence

D'après ces résultats nous remarquons que les performances des 3 configurations sont très proches avec un léger avantage pour la combinaison de l'influence et de l'expertise sur le modèle basé uniquement sur l'influence ou l'expertise.

5.3. Évaluation de l'efficacité du modèle de recherche d'information sociale

Nous comparons dans la figure 5 les valeurs de MAP obtenues par la composante thématique de notre modèle avec $\alpha = 1$ et les 3 configurations sociales avec la bonne valeur du paramètre $\alpha = 0.9$ obtenue par expérimentations. Nous notons que le bon résultat est obtenu par un notre modèle avec la combinaison de la pertinence thématique, de l'expertise et de l'influence dépassant ainsi le modèle de référence basé uniquement sur la pertinence thématique.

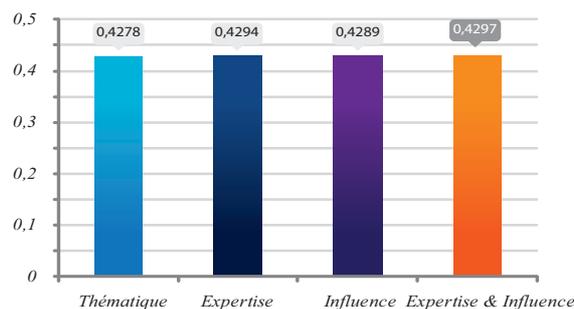


Figure 5 – Comparaison des MAP des différentes configurations

Cependant, nous n'avons pas enregistré dans ces expérimentations une amélioration significative par rapport au modèle de référence. Cette faible performance qui ne dépasse pas 0.5% d'amélioration par rapport au modèle thématique revient à la petite

taille de la collection et au nombre réduit des documents pertinents. Nous expliquons aussi ces résultats par la nature thématique des jugements de pertinence qui se basent sur les 100 premiers documents ordonnés selon le score de pertinence thématique et non sur les interactions sociales.

6. Conclusion

Nous avons proposé dans ce papier un modèle pour la recherche sociale des *tweets*. Ce modèle a la spécificité d'intégrer la pertinence thématique et la pertinence sociale des articles. Cette dernière est estimée par la combinaison de deux scores d'influence et d'expertise de chaque *blogueur*. L'évaluation expérimentale préliminaire que nous avons menée sur une collection d'articles de Twitter montre que cette combinaison permet de mieux évaluer l'importance sociale des *blogueurs*.

En perspective, nous envisageons de mener les expérimentations selon le cadre d'évaluation TREC Microblog qui inclura une collection d'articles et de requêtes de plus grande taille et dont les jugements de pertinence sont de nature sociale. Nous envisageons aussi de comparer les performances de notre modèle avec d'autres modèles sociaux pour la recherche de *tweets*.

7. Bibliographie

- [AGA 08] AGARWAL N., LIU H., TANG L., YU P. S., « Identifying the influential bloggers in a community », *Proceedings of the international conference on Web search and web data mining*, WSDM '08, New York, NY, USA, 2008, ACM, p. 207–218.
- [AME 07] AMER-YAHIA S., BENEDIKT M., BOHANNON P., « Challenges in Searching Online Communities », *IEEE Data Eng. Bull.*, vol. 30, n° 2, 2007, p. 23-31.
- [BAL 08] BALOG K., DE RIJKE M., WEERKAMP W., « Bloggers as experts : feed distillation using expert retrieval models », *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '08, New York, NY, USA, 2008, ACM, p. 753–754.
- [DON 10] DONG A., ZHANG R., KOLARI P., BAI J., DIAZ F., CHANG Y., ZHENG Z., ZHA H., « Time is of the essence : improving recency ranking using Twitter data », *Proceedings of the 19th international conference on World wide web*, WWW '10, New York, NY, USA, 2010, ACM, p. 331–340.
- [EFR 10] EFRON M., « Hashtag retrieval in a microblogging environment », *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '10, New York, NY, USA, 2010, ACM, p. 787–788.
- [Hie 01] HIEMSTRA D., « Using Language Models for Information Retrieval », PhD thesis, Enschede, January 2001.
- [JAV 09] JAVA A., SONG X., FININ T., TSENG B., « Advances in Web Mining and Web Usage Analysis », chapitre Why We Twitter : An Analysis of a Microblogging Community, p. 118–138, Springer-Verlag, Berlin, Heidelberg, 2009.

- [KIR 06] KIRSCH S. M., GNASA M., CREMERS A. B., « Beyond the Web : Retrieval in Social Information Spaces », *In Proceedings of the 28 th European Conference on Information Retrieval (ECIR 2006)*, Springer, 2006.
- [KWA 10] KWAK H., LEE C., PARK H., MOON S., « What is Twitter, a social network or a news media ? », *Proceedings of the 19th international conference on World wide web, WWW '10*, New York, NY, USA, 2010, ACM, p. 591–600.
- [MUT 01] MUTSCHKE P., « Enhancing Information Retrieval in Federated Bibliographic Data Sources Using Author Network Based Stratagems », *Reserach and Advanced Technology for Digital Libraries : 5th European Conference, ECDL 2001, Darmstadt, Germany, September 4-9, 2001 ; Proceedings*, 2001.
- [NOL 09] NOLL M. G., AU YEUNG C.-M., GIBBINS N., MEINEL C., SHADBOLT N., « Telling experts from spammers : expertise ranking in folksonomies », *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval, SIGIR '09*, New York, NY, USA, 2009, ACM, p. 612–619.
- [OUN 10] OUNIS I., MACDONALD C., LIN J., SOBOROFF I., « TREC 2011 Microblog Track », *Text REtrieval Conference TREC*, November 2010.
- [ROB 98] ROBERTSON S. E., WALKER S., HANCOCK-BEAULIEU M., « Okapi at TREC-7 : Automatic Ad Hoc, Filtering, VLC and Interactive », *TREC*, 1998, p. 199-210.
- [SAK 10] SAKAKI T., OKAZAKI M., MATSUO Y., « Earthquake shakes Twitter users : real-time event detection by social sensors », *Proceedings of the 19th international conference on World wide web, WWW '10*, New York, NY, USA, 2010, ACM, p. 851–860.
- [SAN 09] SANKARANARAYANAN J., SAMET H., TEITLER B. E., LIEBERMAN M. D., SPERLING J., « TwitterStand : news in tweets », *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, GIS '09*, New York, NY, USA, 2009, ACM, p. 42–51.
- [TEE 11] TEEVAN J., RAMAGE D., MORRIS M. R., « #TwitterSearch : a comparison of microblog search and web search », *Proceedings of the fourth ACM international conference on Web search and data mining, WSDM '11*, New York, NY, USA, 2011, ACM, p. 35–44.
- [WEN 10] WENG J., LIM E.-P., JIANG J., HE Q., « TwitterRank : finding topic-sensitive influential twitterers », *Proceedings of the third ACM international conference on Web search and data mining, WSDM '10*, New York, NY, USA, 2010, ACM, p. 261–270.
- [YAN 10] YANG Z., GUO J., CAI K., TANG J., LI J., ZHANG L., SU Z., « Understanding retweeting behaviors in social networks », *Proceedings of the 19th ACM international conference on Information and knowledge management, CIKM '10*, New York, NY, USA, 2010, ACM, p. 1633–1636.
- [ZAI 10] ZAIHAN YANG L. H., DAVISON B. D., « Topic-driven Multi-type Citation Network Analysis », *9th International Conference on Adaptivity, Personalization and Fusion of Heterogeneous Information (RIAO), Paris, France, 28/04/2010-30/04/2010*, April 2010.
- [ZHA 07] ZHANG J., TANG J., LI J., « Expert Finding in a Social Network », KOTAGIRI R., KRISHNA P., MOHANIA M., NANTAJEEWARAWAT E., Eds., *Advances in Databases : Concepts, Systems and Applications*, vol. 4443 de *Lecture Notes in Computer Science*, p. 1066-1069, Springer Berlin / Heidelberg, 2007.