

Nouvelle méthode de calcul de la réputation dans les forums de santé

Amine Abdaoui*

*860 rue de St Priest 34095 Montpellier cedex 5
abdaoui@lirmm.fr
www.lirmm.fr/~abdaoui

Résumé. De plus en plus de forums, tels que *Slashdot* ou *Stack Exchange*, proposent des systèmes de réputations qui se basent sur le vote collaboratif. Les utilisateurs peuvent ainsi donner un score à chaque message posté selon sa pertinence ou son utilité. Cependant, ces fonctionnalités de vote sont rarement utilisées dans de nombreuses communautés en ligne tels que les forums de santé. Dans ces forums, les utilisateurs préfèrent poster un nouveau message exprimant de l'accord ou du remerciement vis à vis des messages pertinents plutôt que de cliquer sur un bouton de vote. Dans ce travail, nous proposons d'utiliser ces formes implicites d'expression de la confiance pour estimer la réputation des utilisateurs dans les forums de santé.

1 Introduction

Les forums en ligne sont des espaces d'échanges générés par leurs propres utilisateurs. La qualité des informations postées dépend fortement de l'expertise des auteurs. Avec la croissance rapide des forums et leur large utilisation, il devient de plus en plus difficile de modérer les messages et les utilisateurs manuellement. En effet, de plus en plus de forums proposent des systèmes de réputation automatique pour inférer la confiance que l'on peut accorder à chaque utilisateur et donc aux messages qu'il poste. Si la plupart des forums se basent seulement sur le nombre de messages postés pour calculer la réputation, certains vont plus loin en exploitant l'intelligence collaborative des utilisateurs. Ces derniers donnent des scores aux messages. La réputation d'un utilisateur sera par la suite calculée en agrégeant les scores des messages qu'il a posté (Lampe et Resnick, 2004). Cette idée est implémentée par le groupe d'informations *Slashdot*¹ ou encore le forum de questions-réponses *Stack Exchange*². Cependant, le vote collaboratif n'est pas si populaire dans d'autres communautés en ligne comme les forums de santé. En effet, les utilisateurs de ses forums préfèrent remercier un auteur en postant un nouveau message dans lequel ils expriment textuellement leur gratitude plutôt que cliquer sur un bouton pour voter. L'objectif de ce travail est d'utiliser ces expressions textuelles de la confiance pour inférer la réputation des utilisateurs des forums de santé.

Beaucoup de définitions de la confiance et de la réputation existent dans la littérature (Golbeck, 2009). Ici, nous définissons la confiance qu'accorde un utilisateur *A* à un utilisateur *B*

1. www slashdot.org

2. www stackexchange.com

comme : *la croyance de A dans la véracité des informations postées par B* et la réputation de l'utilisateur B comme : *l'agrégation des valeurs de confiance accordées à l'utilisateur B*. Pour calculer une telle mesure de réputation sur la base du contenu textuel, nous devons avant tout connaître le(s) destinataire(s) de chaque message. Cependant, certains forums de santé ne permettent pas de répondre à un message spécifique mais à tout le sujet de discussion. Et même lorsque la réponse directe est implémentée, elle demeure très peu utilisée. Pour cela, nous proposons une heuristique à base de règles pour extraire un réseau des réponses entre utilisateurs. Ensuite, nous évaluons la confiance exprimée dans le contenu de chaque message. Nous nous intéressons particulièrement aux expressions d'accord, de désaccord et de remerciement. Enfin nous proposons une métrique pour agréger la réputation de chaque utilisateur sur la base des réponses reçues tout en prenant en considération les aspects de propagation de la confiance. Cet article présente l'extension de (Abdaoui et al., 2015) en appliquant la méthode à un deuxième forum et en étendant l'ensemble d'annotations manuelles.

2 État de l'art

La plupart des méthodes d'extraction des réseaux d'utilisateurs dans les communautés en ligne (première étape de notre méthode) se basent sur la structure HTML de la page web (Adamic et al., 2008; Stavrianou et al., 2009). Très peu utilisent le contenu textuel des messages pour identifier leurs destinataires. Parmi ces travaux, (Gruzd et Haythornthwaite, 2008) ont proposé un système de reconnaissance d'Entités Nommées pour l'extraction des mentions de noms dans le contenu textuel des messages. Ils ont utilisé un dictionnaire de noms avec des règles définies manuellement. (Forestier et al., 2011) ont proposé d'extraire trois types d'interactions entre utilisateurs : les citations directes, les citations de noms et les citations de textes. Si la première catégorie de citation peut être déduite via la structure de la page, les deux autres nécessitent l'analyse du contenu des messages. Les auteurs ont testé plusieurs types de distances lexicales pour extraire ces deux derniers types de citations.

De façon générale, les méthodes de calcul de la confiance et de la réputation dans les forums en ligne peuvent être organisées en deux catégories. La première catégorie définit un certain nombre de métriques qui se basent sur la structure du forum telle que le nombre de messages postés, la distance entre les messages, le nombre de citations, etc. Une telle méthode a été proposée par (Skopik et al., 2009). La deuxième catégorie utilise le contenu textuel des messages pour estimer la confiance que l'on peut accorder à un message et donc par agrégation les réputations de leurs auteurs. Par exemple, (Wanas et al., 2008) se sont inspirés de *Slashdot* pour modéliser les raisons pour lesquels les utilisateurs trouvent qu'un message est pertinent ou non. À l'inverse, nous nous intéressons ici aux messages qui sont destinés à chaque utilisateur au lieu des messages que lui-même a postés. Nous nous intéressons particulièrement aux expressions d'accord, de désaccord et de remerciements.

Enfin, un effort important a été réalisé pour mesurer l'autorité des pages web (Page et al., 1999) en incluant les aspects de propagation de cette autorité. Nous nous inspirons de ces travaux pour proposer une métrique de calcul de la réputation qui donne plus de valeur aux messages postés par des utilisateurs dignes de confiance.

3 Matériels et méthodes

Dans cette section nous décrivons les corpus utilisés et les méthodes proposées.

Corpus : Deux forums de santé Français ont été récupérés. Le premier est spécialisé dans le cancer du sein regroupant (à la date de sa récupération) 675 utilisateurs, 1 050 sujets de discussion et 16 961 messages. Le deuxième traite des maladies liées à la thyroïde et contenait au moment de sa récupération 13 803 utilisateurs, 37 857 sujets et 269 073 messages.

Formulation mathématique : Soit $G = (V, E, t, r)$ un multigraphe où : V représente l'ensemble d'utilisateurs (nœuds), E l'ensemble des réponses (arêtes), t une fonction qui retourne l'émetteur du message et r une fonction qui retourne le destinataire du message :

$$t : \begin{array}{l} E \longrightarrow V \\ e \longmapsto t(e) \end{array} \quad r : \begin{array}{l} E \longrightarrow V \\ e \longmapsto r(e) \end{array}$$

Soit $v \in V$ un utilisateur, alors $E_v \subseteq E$ est l'ensemble des messages qui répondent à v :

$$E_v = \{e \in E : r(e) = v\}$$

Soit E_v^+ , E_v^- , E_v^n les sous-ensembles de réponses de E_v exprimant la confiance, la méfiance et le neutre respectivement. À noter que : $E_v^+ \cup E_v^- \cup E_v^n = E_v$ et $E_v^+ \cap E_v^- \cap E_v^n = \emptyset$.

Extraction du multigraphe : Pour chaque message (à l'exception du premier message de chaque sujet de discussion), nous vérifions les règles suivantes séquentiellement :

1. Les citations structurelles directes sont reliées à l'auteur du message cité.
2. Le deuxième message de chaque sujet de discussion est relié à l'activateur³.
3. Les messages contenant des mentions de noms ou de pseudonymes d'utilisateurs ayant postés avant dans le même sujet sont reliés à ces utilisateurs. Plusieurs prétraitements ont été appliqués : mise en minuscules, remplacement des accents, des caractères spéciaux, etc.
4. Les messages contenant des marqueurs de groupe ('*Bonjour à toutes*', etc.) sont reliés aux trois derniers utilisateurs ayant postés dans le sujet.
5. Les messages contenant du tutoiement sont reliés à l'auteur du message précédent.
6. Les messages suivant de l'activateur sont reliés aux utilisateurs ayant postés après son dernier message dans le sujet (limité à trois utilisateurs maximums).
7. Les messages contenant une question sont reliés aux trois derniers utilisateurs ayant postés dans le même sujet.
8. Si une question a été posée dans le même sujet alors le message répond à son auteur.
9. Si aucune des règles ci-dessous n'est applicable alors chaque message est relié à l'auteur du message précédent.

3. L'utilisateur ayant ouvert le sujet en postant le premier message

Analyse textuelle des messages : Afin d'identifier les messages exprimant de la confiance (accord et remerciements) et ceux exprimant la méfiance (désaccord), nous avons créé manuellement deux listes de termes (confiance : 33 termes, méfiance : 15 termes). Ces listes ont été obtenues d'abord par des annotations manuelles sur l'outil *Brat*⁴. Les annotateurs devaient décider si un message exprimait de la confiance, de la méfiance ou aucun des deux et sélectionner le(s) terme(s) justifiant leur choix. Par la suite, ces listes ont été prétraitées (mises en minuscules, lemmatisées et corrigées⁵) et enrichies. Les termes de ces listes ont été recherchés à l'intérieur des messages prétraités. La valeur de confiance d'un terme est inversée s'il se trouve sous la portée d'un négateur. Enfin, chaque message a été mis dans la classe majoritaire de ces termes (confiance, méfiance, neutre).

Métriques proposées : Pour chaque utilisateur v , nous définissons sa réputation $R(v)$ de la manière suivante :

$$R_{n+1}(v) = \begin{cases} \frac{\sum_{e \in E_v^+} R_n(t(e))}{\sum_{e \in E_v^+} R_n(t(e)) + \sum_{e \in E_v^-} R_n(t(e))} & , \text{ Si } E_v^n \neq E_v \\ 0,5 & , \text{ Sinon.} \end{cases}$$

Cette équation est récursive, elle peut être calculée en commençant avec des réputations égales à 1 et en répétant jusqu'à convergence. Cette métrique dépend du nombre de messages reçus de confiance et de méfiance, mais aussi de la réputation de leurs auteurs. Nous définissons deux métriques complémentaires : le pourcentage de message neutres $NR(v)$ et la fiabilité de la réputation calculée $Rel(R(v))$:

$$NR(v) = \begin{cases} \frac{|E_v^n|}{|E_v|} & , \text{ Si } E_v \neq \emptyset \\ 0 & , \text{ Sinon.} \end{cases} \quad Rel(R(v)) = \begin{cases} \frac{|E_v|}{maxR} & , \text{ Si } |E_v| < maxR \\ 1 & , \text{ Sinon.} \end{cases}$$

$maxR$ est une constante qui représente le nombre minimum de réponses un utilisateur doit recevoir pour être fiable à 100%.

4 Résultats

Afin d'évaluer nos méthodes, nous avons annoté manuellement 20 sujets de discussions à partir de chaque forum. Chaque sujet a été annoté par 3 annotateurs différents. Le but était de trouver manuellement le(s) destinataire(s) de chaque message et de décider la confiance qu'il exprime. Les annotations retrouvées par au moins deux annotateurs ont été validées (vote majoritaire). Le tableau 1 présente les résultats obtenus par l'heuristique d'extraction du multigraphe. Nous nous comparons avec une baseline considère que chaque message répond à l'activateur (Activateur) et une autre qui considère que chaque message répond au précédent (Précédent). Notre heuristique a obtenu les meilleurs F-mesures sur les deux forums.

Le tableau 2 montre les moyennes pondérées des précisions, rappels et F-mesures du système de classification des messages selon la confiance exprimée pour les deux forums. Les résultats obtenus sont autour de 0.80 sur les deux forums.

4. <http://www.brat.nlplab.org/>

5. <http://www.aspell.net>

Forum	Méthode	Précision	Rappel	F-mesure
CancerDuSein.org	Activateur	0,516	0,330	0,402
	Précédent	0,650	0,416	0,507
	Notre méthode	0,807	0,663	0,728
Forum-Thyroides.net	Activateur	0,455	0,411	0,431
	Précédent	0,793	0,716	0,753
	Notre méthode	0,714	0,821	0,764

TAB. 1 – Évaluation de l’heuristique d’extraction des liens sur les deux forums

Forum	Précision	Rappel	F-mesure
CancerDuSein.org	0,86	0,84	0,84
Forum-Thyroide.net	0,82	0,79	0,81

TAB. 2 – Évaluation de la classification des messages sur les deux forums

Enfin, la figure 1 montre les réputations des utilisateurs de chaque forum pour lesquels la fiabilité est supérieur à 0,5 en fonction du nombre de messages postés. $maxR$ a été fixé à la moyenne des réponses reçues par les utilisateurs de chaque forum. Globalement, les réputations calculées sur les deux forums sont bonnes. Ceci peut être expliqué par la nature des discussions sur les forums de santé (aide, support émotionnel, etc.). En effet très peu de désaccord est exprimé. Afin de valider les réputations calculées, les modérateurs du deuxième forum nous ont fournis les pseudonymes de 8 utilisateurs qui doivent avoir d’excellentes réputations. Ces derniers figurent parmi les 60 utilisateurs ayant les meilleurs réputations selon notre méthode.

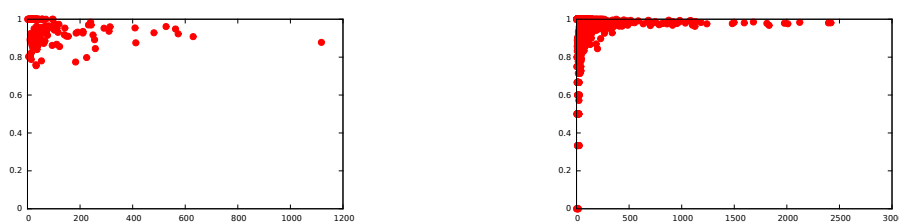


FIG. 1 – Réputations en fonction du nombre de messages postés pour chaque forum

5 Perspectives

Plusieurs perspectives peuvent être considérées afin d’améliorer ce travail. D’abord, nous allons tester notre méthode sur d’autres forums non liés à la santé. Ensuite, nous voulons élargir l’ensemble annoté manuellement pour l’évaluation de nos méthodes en faisant appel aux outils de crowdsourcing. De plus, nous prévoyons d’appliquer les méthodes de calcul de l’autorité des pages Web (PageRank, HITS, etc.) sur le multigraphe des réponses entre utilisateurs

extrait en utilisant notre heuristique. En effet, ces derniers ont été appliquées efficacement pour mesurer l'autorité des utilisateurs de groupes Facebook ou encore les adhérents d'une association scientifique. Enfin, une évaluation experte des réputations calculées sera effectuée avec la participation des modérateurs des deux forums.

Références

- Abdaoui, A., J. Azé, S. Bringay, et P. Poncelet (2015). Collaborative content-based method for estimating user reputation in online forums. In *Proceedings of the 16th International Conference on Web Information Systems Engineering–WISE 2015*, pp. 292–299. Springer.
- Adamic, L. A., J. Zhang, E. Bakshy, et M. S. Ackerman (2008). Knowledge sharing and yahoo answers : everyone knows something. In *WWW 2008*, pp. 665–674. ACM.
- Forestier, M., J. Velcin, et D. Zighed (2011). Extracting social networks enriched by using text. In *ISMIS 2011*, pp. 140–145. Springer.
- Golbeck, J. (2009). Trust and nuanced profile similarity in online social networks. *ACM Transactions on the Web (TWEB)* 3(4), 12.
- Gruzd, A. A. et C. Haythornthwaite (2008). Automated discovery and analysis of social networks from threaded discussions. University of Illinois at Urbana-Champaign.
- Lampe, C. et P. Resnick (2004). Slash (dot) and burn : distributed moderation in a large online conversation space. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 543–550. ACM.
- Page, L., S. Brin, R. Motwani, et T. Winograd (1999). The pagerank citation ranking : bringing order to the web. Stanford InfoLab.
- Skopik, F., H.-L. Truong, et S. Dustdar (2009). Trust and reputation mining in professional virtual communities. In *ICWE 2009*, pp. 76–90. Springer.
- Stavrianou, A., J. Velcin, et J.-H. Chauchat (2009). Definition and measures of an opinion model for mining forums. In *International Conference on Advances in Social Network Analysis and Mining, 2009. ASONAM'09.*, pp. 188–193. IEEE.
- Wanas, N., M. El-Saban, H. Ashour, et W. Ammar (2008). Automatic scoring of online discussion posts. In *the 2nd ACM Workshop on Information Credibility on the Web*, pp. 19–26. ACM.

Summary

More and more forums, such as *Slashdot* ou *Stack Exchange*, offer reputation systems based on collaborative votes of posted messages. Indeed, users can give a score to each message according to its relevance or usefulness. However, these voting functionalities are not used in many online communities such as health forums. In these forums, users prefer posting a new message (reply) where they express agreement or thanking about relevant posts rather than pressing a vote up button. In this work, we suggest to use these hidden expressions of trust in order to estimate user reputations in online forums.