

Détection des profils à long terme et à court terme dans les réseaux sociaux

Dieudonné Tchunte, Marie Françoise Canut, Nadine Baptiste-Jessel

Institut de Recherche en Informatique de Toulouse (IRIT)
118, Route de narbonne, 31062, Toulouse, France.
Dieudonne.Tchunte@irit.fr, canut@iut-blagnac.fr, baptiste@irit.fr

Résumé

La conception des profils et contextes utilisateurs se situe au cœur de l'étude et de la mise en œuvre des mécanismes de personnalisation ou d'adaptation de contenus (recherche d'information, systèmes de recommandation, etc.). Plusieurs modèles et dimensions de profils et contextes sont décrits dans la littérature. Dans la vie réelle tout comme dans les systèmes d'information, le comportement de l'utilisateur est très souvent influencé par son environnement social. Cependant, la dimension sociale des profils et contextes utilisateurs reste très peu étudiée et évaluée. Dans cet article, nous présentons une méthode de visualisation des profils utilisateurs permettant d'évaluer la pertinence du réseau social de l'utilisateur dans l'évolution de son profil. L'expérimentation de la méthode à partir de Facebook permet d'identifier d'une part, les centres d'intérêts à court-terme et à long-terme des profils utilisateurs, et d'autre part, l'influence réelle à court-terme et à long-terme du réseau social de chaque utilisateur. Ces résultats démontrent l'intérêt de modéliser et d'intégrer une dimension sociale dans les profils et contextes utilisateurs, afin de tenter d'améliorer les mécanismes de personnalisation ou d'adaptation de contenus.

1 Introduction

La conception et la mise en œuvre des profils utilisateurs est une problématique développée dans les systèmes de personnalisation ou d'adaptation de contenus [Boughanem,10]. Dans ces systèmes, il est indispensable d'associer le contexte de l'utilisateur à son profil indépendamment de la méthode de représentation des profils. En considérant le contexte temporel par exemple, pour un même utilisateur, le centre d'intérêt « Java » peut correspondre à certaines périodes à un intérêt pour le langage de programmation Java, mais à d'autres périodes à un intérêt pour du tourisme (île Indonésienne Java). Dans le cas de l'exemple précédent, en plus des analyses contextuelles, il serait également nécessaire de considérer des analyses sémantiques de profils afin de pouvoir distinguer le sens du centre d'intérêt « Java » (programmation Java ou île Java). Ainsi, dans la littérature, on distingue : (i) des travaux orientés vers la représentation des éléments de profils (vecteurs de mots, réseaux sémantiques, réseaux de concepts) [Gauch et al., 07] ; (ii) des travaux orientés vers la sémantique des profils [Chevalier et al., 07] ; (iii) des travaux orientés vers les profils contextuels [Schilit et al., 94] ; (iv) des travaux orientés vers les profils sémantiques et contextuels [Abbar et al., 09] ou des profils multidimensionnels [Amato et al., 99].

Cependant, tous ces travaux actuels de conception ou de représentation des profils et contextes utilisateurs modélisent les utilisateurs de manière individuelle. Pourtant, dans la vie réelle, les comportements des utilisateurs sont fortement influencés par leurs interactions

sociales ou leur environnement social [Granovetter et al., 73]. L'influence sociale est d'autant plus importante qu'elle est à la base de domaines émergents tels que le marketing viral, les systèmes de recommandation sociaux, ou la recherche d'information sociale [Jabeur, Tamine, 09], qui s'appuient en général sur les profils utilisateurs. Certains travaux s'intéressent déjà à cette problématique et proposent des mécanismes adaptatifs intégrant des données sociales. C'est le cas par exemple de l'expérimentation menée par [Lefevre, Cabanac, 10] qui propose un système de recommandation thématique-sociale d'articles bibliographiques. [Fogelman, 10] s'intéresse aussi à l'usage des données sociales dans le processus de modélisation en data mining, dans le but de pallier l'imprécision des modèles construits par les algorithmes de data mining, qui supposent généralement une indépendance totale entre les variables étudiées. Dans les deux derniers travaux cités, les techniques utilisées sont pertinentes, cependant elles restent très spécifiques à des domaines d'applications (recherche d'information bibliographique, data mining, etc.). Il conviendrait plutôt de concevoir des profils utilisateurs génériques prenant en compte les relations et les interactions sociales, afin qu'ils puissent être réutilisés dans plusieurs domaines d'applications (marketing viral, recherche d'information, système de recommandation, data mining, etc.). Le premier enjeu pour ce type de modélisation est de pouvoir capturer et juger réellement de la pertinence (l'influence) du réseau social de l'utilisateur dans son profil. Ainsi, dans cet article, nous présentons une méthode de conception et de visualisation de l'influence du réseau social d'un utilisateur sur son profil.

Dans la section qui suit, nous présentons les différentes étapes de la méthodologie proposée. Ensuite, nous présentons l'expérimentation réalisée à partir de profils extraits de la plateforme Facebook. Enfin, nous clôturons par les conclusions et les perspectives de ce travail.

2 Méthodologie

L'intérêt de la méthode proposée est de pouvoir intégrer et juger de la pertinence de l'intégration d'une dimension sociale dans le développement de profils utilisateurs. Cette méthode doit également pouvoir prendre en considération les meilleurs critères de conception de profils actuellement présents dans la littérature, à savoir : la séparation des éléments du profil et des contextes (pour optimiser la qualité des résultats issus des mécanismes utilisant ces profils), la prise en compte de la sémantique (pour éviter toute sorte d'ambiguïté sur la terminologie manipulée). Afin de répondre à ces contraintes et de pouvoir visualiser graphiquement les résultats, nous réutilisons la méthodologie de la plateforme *Tétralogie*¹ (figure 1, page suivante) dans laquelle nous ajoutons des étapes pour construire et prendre en compte le réseau social des utilisateurs.

La méthodologie de *Tétralogie* consiste à extraire et visualiser des unités sémantiques pertinentes et leurs relations par fouille de corpus de données textuelles indépendamment de leur nature (structurées, semi-structurées, non structurées) et de leur source (bases de données, CD/DVD, Web, etc.) [Dousset, 03]. Nous nous sommes appuyés sur cette méthode pour construire les profils utilisateurs par extraction des unités sémantiques pertinentes à partir des traces d'activités des utilisateurs dans un système d'information. La méthode offre déjà l'avantage d'implémenter plusieurs techniques permettant d'associer une sémantique forte aux analyses (filtres d'unités sémantiques, dictionnaires, thésaurus et ontologies)

¹ <http://atlas.irit.fr/>

[Dousset, 03], et de visualiser des réseaux sémantiques construits à partir de matrices de cooccurrences (2D) d'unités sémantiques.

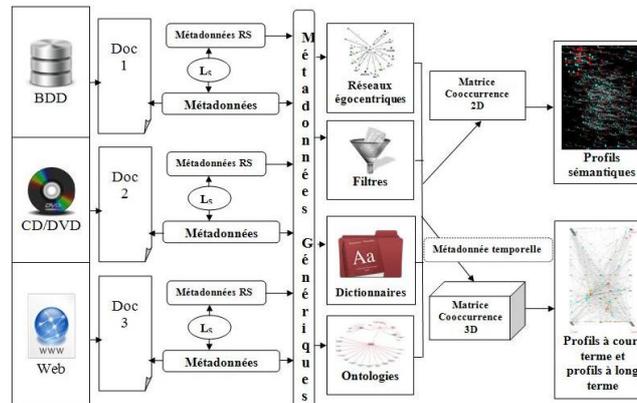


FIG.1 - la méthodologie proposée pour la conception de profils utilisateurs dans les réseaux sociaux.

Ces dernières matrices peuvent prendre en compte des dimensions supplémentaires (qui représentent dans notre étude les contextes des profils) telle que la dimension temporelle [Gay, Loubier, 09].

Pour pouvoir considérer l'influence du réseau social d'un utilisateur dans son profil, nous construisons dans un premier temps un graphe $G = (V, E)$ dans lequel V représente l'ensemble des utilisateurs extraits du corpus, et E représente l'ensemble des liens entre ces utilisateurs. La construction du réseau social consiste à définir (avec un expert du domaine étudié), la sémantique du lien permettant d'ajouter un couple de nœuds de V , dans E . La définition de ce lien sémantique est réalisée ici, par l'ajout d'une fonction nommée Ls^2 qui définit la condition à respecter par les valeurs d'une ou plusieurs métadonnées afin de créer ou non un lien entre chaque couple d'utilisateurs de V dans E . De la fonction Ls , nous construisons alors de nouvelles métadonnées décrivant comment générer le réseau social (métadonnées RS). La fonction Ls pourrait, par exemple, indiquer la création d'un lien entre deux clients d'une société de télécommunications s'ils se sont échangés au moins 10 SMS les weekends, ou encore la création d'un lien entre deux utilisateurs de Facebook s'ils sont amis, etc.

Une fois le réseau social spécifié, le graphe déduit est utilisé pour enrichir les analyses de chaque profil. Etant donné que nous nous intéressons aux utilisateurs qui influencent directement chaque profil analysé, nous adoptons une approche d'analyse *égocentrée* (sur les relations directes qu'entretient une personne avec d'autres (alters) et les relations que ces alters ont entre eux). Ceci à la différence des approches *sociocentrées* (qui considèrent à chaque fois le graphe entier dans les analyses) telles que celle utilisée dans [Fogelman, 10], supposent d'avoir accès à la totalité des réseaux (ce qui n'est très souvent pas le cas) et deviennent très complexes à mettre en œuvre pour des graphes de très grande taille (complexité de calcul des mesures de centralité par exemple). Une approche *sociocentrée* aurait donc des conséquences lourdes sur les mécanismes adaptatifs qui nécessitent des profils constamment

mis à jour et des réponses en temps réel. L'approche égocentrée utilisée ici, privilégiera donc uniquement les utilisateurs ayant des liens forts avec chaque profil analysé [Granovetter, 73] au détriment des liens faibles (qui nécessitent d'analyser un réseau en entier). Cet avantage correspond plus à nos besoins car nous nous intéressons aux liens d'influences forts entre profils, plutôt qu'à la facilité de diffusion de l'information dans un réseau (liens faibles).

Afin de mieux appréhender l'influence du réseau égocentrique d'un utilisateur dans son profil, nous avons choisi de privilégier les analyses suivant le contexte temporel afin de pouvoir distinguer les utilisateurs qui auraient des influences constantes sur le profil analysé, des utilisateurs qui influencent périodiques ou temporaires. Nous nous appuyons dans cet article sur des matrices temporelles 3D et sur l'algorithme de *Visugraph* [Gay, Loubier, 09] qui permet de visualiser à un niveau de granularité très élevé, des volumes important de données multidimensionnelles. L'expérimentation réalisée est présentée dans la section qui suit.

3 Expérimentation

L'expérimentation de la méthode a été réalisée sur la plateforme *Facebook* qui dispose d'une API permettant à des développeurs tiers d'extraire des informations sur les éléments de profil renseignés par les utilisateurs et sur un nombre relativement important de leurs traces d'activités. De plus l'API permet également d'extraire intégralement le réseau égocentrique de l'utilisateur qui l'installe. Ainsi, nous avons développé une application *Facebook*³ dédiée à des utilisateurs volontaires qui nous ont donné le droit d'accéder à leurs données durant toute la durée de notre étude. L'application a été installée par 81 utilisateurs volontaires, et nous avons accédés à 7081 profils par le biais de requêtes *FQL* (Facebook Query Language) d'extraction de réseaux égocentriques. Les traces d'activités extraites sont principalement les éléments publiés par les utilisateurs (statuts, articles, tags, URL, photos, etc.). Pour chaque élément publié, deux types de métadonnées sont créées: les métadonnées de contenus (utilisateurs impliqués dans la publication, et le contenu textuel publié), et les métadonnées de contexte (le contexte temporel dans notre cas). La fonction *Ls* considérée ici, est la relation d'amitié entre les profils ayant interagi au moins une fois. Un exemple de graphe construit après croisement des unités sémantiques extraites du réseau égocentrique et du contexte temporel est présenté sur la figure 2 (page suivante).

Cette figure présente le sous-graphe restreint au profil de l'utilisateur le plus actif de l'échantillon de données (utilisateur nommé Up ici). En fonction de la fréquence des activités, le temps a été divisé en quatre périodes (semestres). Les centres d'intérêts (respectivement les utilisateurs) dont les poids (respectivement les fréquences d'activités) sont quasi constants durant toutes les périodes, se rapprochent du centre du graphe et représentent les centres d'intérêts à long-terme de l'utilisateur (respectivement les amis de l'utilisateur dont les interactions avec l'utilisateur sont constantes dans le temps 'influence à long-terme'). Sur le graphe (cadre en vert, figure 2), on peut ainsi identifier les centres d'intérêt à long-terme (photos, article, champagne, vidéos, musique, jeu, etc.) pour lesquels seul l'utilisateur (U29) est influent dans le réseau égocentrique de « Up ». Les centres d'intérêt (respectivement les utilisateurs) dont le poids sur une seule période (respectivement la fréquence d'activités sur une seule période) est très élevé relativement aux poids sur les autres périodes (respective-

³ Lien vers l'application développée : <http://apps.facebook.com/analyze-network>

ment aux fréquences d'activités sur les autres périodes), se rapprochent de la période considérée et représentent les centres d'intérêt à court-terme de l'utilisateur (respectivement les amis de l'utilisateur qui sont actifs ou qui l'influencent uniquement sur la période considérée – 'influence à court-terme').

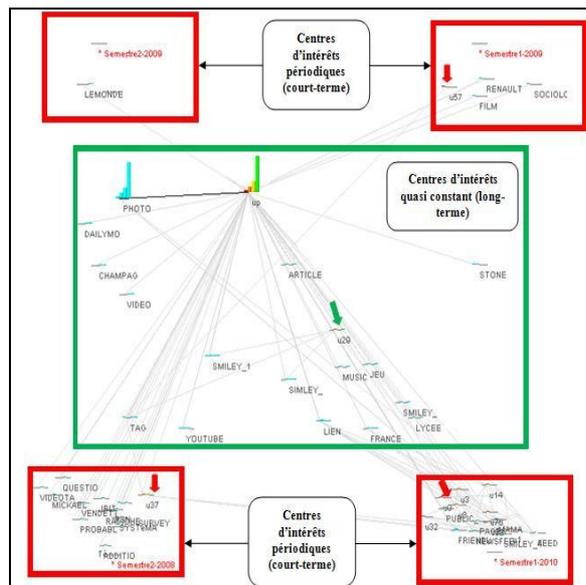


FIG.2 – visualisation de profil à court-terme et à long-terme

Sur le graphe (cadres en rouge, figure 2), on peut ainsi identifier les centres d'intérêts (ou utilisateurs actifs) à court-terme du semestre 1 de 2009 (films, Renault, sociologie, « U57 »), du semestre 1 de 2010 (« U3 », « U14 », « U32 », etc.), du semestre 2 de 2008 (Michael Vendetta, « U37 », etc.), du semestre 2 de 2009 (le site web *lemonde*, centre d'intérêt exclusivement propre à l'utilisateur).

4 Conclusion et perspectives

Dans cet article, nous avons proposé et expérimenté une méthode de visualisation des profils utilisateurs à long-terme et à court-terme, en prenant en compte l'influence du réseau social de l'utilisateur. Afin de réduire la complexité en temps de calcul et de pallier le manque récurrent de données sur la totalité des réseaux sociaux, nous optons pour des analyses égocentrées de profils utilisateurs. Ces analyses, comme nous avons pu le démontrer dans notre expérimentation, permettent de bien mettre en évidence les « liens forts » d'influence du réseau égocentrique d'un utilisateur dans son profil. Ces résultats motivent des perspectives d'intégration d'une dimension sociale dans les modèles de profils utilisateurs, afin de tenter d'améliorer les mécanismes nécessitant des données sociales telles que les systèmes de recommandation sociaux, la recherche d'information sociale, etc.

Bibliographie

- Abbar S., Mokrane B., Dimitre K., Stéphane L. (2009). *A contextualization service for a Personalized Access Model*. EGC 2009: 265-270.
- Amato Giuseppe, Umberto Straccia: *User Profile Modeling and Applications to Digital Libraries (199)*. ECDL 1999: pp184-197.
- Boughanem M. *Recherche d'information contextuelle (2010)*. Actes de l'école d'été Web intelligence 2010, Web centré sur l'utilisateur, Saint-Germain-Au-Mont-d'Or, France. pp 47-72.
- Chevalier M., Julien C., Soulé-Dupuy C., Vallés-Parlangeau N. (2007). *Personalized Information Access Through Flexible and Interoperable Profiles*. WISE-PAWI 2007, Nancy-France, Vol. LNCS 4832, p. 374-385.
- Dousset Bernard (2003). *Intégration de méthodes interactives de découverte de connaissances pour la veille stratégique*. Habilitation à diriger des recherches, Université Paul Sabatier, novembre 2003.
- Everett Martin, Stephen P. Borgatti (2005). *Ego network betweenness (2005)*. In Social Networks, Elsevier, Vol. 27, No. 1., pp. 31-38.
- Fogelman F. S. (2010). *Mise en œuvre industrielle d'un composant de data mining sur les réseaux sociaux. Quelques exemples d'application*. Journée Réseaux sociaux analyse et data mining, SFDS, Paris, 16 Février 2010.
- Gauch Susan, Mirco S., Aravind C., Alessandro M. (2007). *User profiles for Personalized Information Access*. in The Adaptive Web, Vol. 4321 (2007), pp. 54-89
- Gay Brigitte, Eloïse Loubier (2009). *Dynamics and Evolution Patterns of Business Networks*. In 2009 Advances in Social Network Analysis and Mining, pp. 290-295.
- Granovetter M.S. (1973). *The Strength of Weak Ties*. The American Journal of Sociology, Vol. 78. No. 6, May 1973, pp. 1360-1380.
- Jabeur Lamjed B., Lynda Tamine (2010). *Vers un modèle de Recherche d'Information Sociale pour l'accès aux ressources bibliographiques*, Conférence francophone en Recherche d'Information et Applications CORIA 2010.
- Lefevre A., Cabanac G. (2010), *Confrontation à la perception humaine de mesures de similarités entre membre d'un réseau social académique*. MARAMI 2010, Toulouse, octobre 2010.
- Schilit Bill N., Norman A., and Roy W. (2004). *Context-Aware Computing Applications*. IEEE Workshop on Mobile Computing Systems and Applications (WMCSA'94), Santa Cruz, CA, US. pp. 89-101.

Summary

User profile modeling is at the heart of contents adaptative systems like personalized system or recommender system. Existing works in user profile modeling usually combine contextual and/or semantic information's to user profiles elements for improving the relevance of personalized or recommended contents. However, these models are not really realistic because the user is always modeled as an independent entity; whereas in reality, the user behavior is strongly influenced by its social ties and social interactions. This social dimension of user profiles is less studied and evaluated in the literature. In this paper, we present and experiment (on Facebook) a method for visualizing the evolution of user profiles (short-term and long-term profiles) and for depicting the relevance over time of social networks on each profile. This study aims to demonstrate the interest to model the social network of the user in his profile.