Approche de recommandation de tags par niveau

Manel Hmimida*

*DICEN, Dispositifs d'Information et de Communication à l'Ère du Numérique
CNAM, Conservatoire National des Arts et Métiers 2, rue de Conté 75003.
manel.hmimida@cnam.fr

Résumé. Dans cet article, nous proposons une nouvelle approche de recommandation de tags dans des folksonomies. Notre approche est fondée sur la structuration des tags déjà utilisés par les utilisateurs pour annoter des ressources dans des clusters. La recommandation intervient à deux niveaux : la recommandation de clusters puis la recommandation de tags. Pour rendre les recommandations plus pertinentes, notre approche fait intervenir l'utilisateur pour la validation de recommandation de clusters proposés en premier temps par le système. Une fois que l'utilisateur a fait son choix de cluster, le système lui propose des recommandations de tags.

1 Introduction

Avec l'essor du Web, les sites de partage social de ressources appelés aussi folksonomies sont devenus populaires. Ces systèmes permettent aux utilisateurs d'annoter leurs ressources via des mots clés (tags). Ces tags permettent de décrire et de catégoriser les ressources pour des fins organisationnelles et de partage.

Les folksonomies se distinguent par le type de ressource qu'elles supportent. Parmi les systèmes les plus connu : Flickr permettant le partage des photos, CiteULike et Connotea le partage des références bibliographiques. Le but de tags est de faciliter et d'accélérer les tâches de navigation et de recherche. Ces tags sont crées librement par les utilisateurs. Le même tag peut être utilisé différemment par différents utilisateurs. Et une même ressources peut être annotée par différents tags. Par conséquence, nous nous retrouvons face au problème d'ambiguïté de tags. Pour contourner cette problématique, la plupart de ces systèmes disposent d'un mécanisme de recommandation de tags permettant de réduire l'effort cognitif de l'utilisateur et de lui faciliter le bon repérage de bons tags pour annoter ses ressources.

Dans ce papier, nous proposons un nouveau modèle de recommandation dit par niveau afin d'assister l'utilisateur dans le processus de choix de tags de ses ressources. Le mécanisme de recommandation consiste à utiliser les traces d'indexation individuelles de ressources effectuées par les utilisateurs pour apprendre leurs habitudes et leurs préférences. De plus, nous générons la couche clusters pour ajouter un niveau de filtrage, par l'élimination dès le départ de clusters non intéressants. Par conséquent, la quantité de données traitées est nettement moins importante que celle de données de départ. La génération de recommandations s'effectue à deux niveaux. En premier temps le système génère des clusters, et en fonction de celles choisies par l'utilisateur, il recommande en deuxième temps les tags candidats associés à son choix.

Nous proposons un algorithme de recommandation inspiré de l'algorithme de règles d'associations Agrawal et al. (1993). Le même algorithme est utilisé afin de satisfaire deux objectifs : la recommandation de clusters et la recommandation de tags. Cet article s'organise comme suit. La section 2 présente un aperçu sur l'état de l'art dans le domaine des systèmes de recommandation : les principes et les techniques de recommandation existantes. Nous introduisons dans la section 3 le principe de notre approche. Nous finissons par conclure et présenter les perspectives.

2 Etat de l'art

De nombreux travaux ont adressé le problème de recommandation de tags dans des folksonomies Xu et al. (2006) Jäschke et al. (2007) Jäschke et al. (2008) Lu et al. (2009). Ces travaux peuvent être divisés en deux classes :

- les approches basées sur le contenu,
- et les approches topologiques.

Parmi les travaux de recommandation basés sur le contenu, nous citons le travail proposé dans Lu et al. (2009). Cette approche de recommandation de tags est appliquée dans les pages Web dont le but est de pouvoir générer des recommandations même sans avoir des informations préalablement connues sur les tags. En plus, leur proposition est motivée par le constat que les pages Web similaires ont tendance à avoir les mêmes tags. Chaque page peut donc partager les tags qu'elles possèdent avec les pages Web similaires. L'avantage des approches basées sur le contenu est de pouvoir préconiser des tags à l'utilisateur même si ces tags n'ont jamais été utilisés avant. L'inconvénient de ce genre d'approches est qu'elles sont incapables de fournir de bonnes recommandations dans le cas où le contenu de ressources est limité.

Dans la catégorie des approches topologiques, nous citons le travail de Jäschke et al. (2008). ici, les auteurs comparent un ensemble des techniques de recommandation comme le filtrage collaboratif, FolkRank. Les auteurs ont évalué ces techniques à grande échelle sur un ensemble de données réelles. Ces expérimentations montrent des résultats plus performants en utilisant des techniques de recommandation dans les systèmes de marquage par rapport aux méthodes non personnalisées. Dans cette même catégorie d'approche, un travail récent de Pujari et Kanawati (2012) proposant une approche de prévision de liens dans les graphes bi-partie en utilisant des approches d'apprentissage automatique supervisé. Pour un utilisateur cible, et une ressource cible, l'algorithme calcule une liste des tags les plus adaptés à annoter la ressource par l'utilisateur cible. Notre approche s'ajoute à la catégorie des approches topologiques dont l'originalité est la recommandation à deux niveaux.

3 Approche proposée

Notre approche consiste à calculer la recommandation à deux niveaux : de plus général au plus détaillé. Vu que la structure d'une folksonomie est définie par un graphe tripartite (utilisateurs, tags, ressources), elle ne dispose pas de la couche cluster, nous sommes obligés de créer ce premier niveau de recommandation (clusters). D'abord, nous commençons par décomposer le graphe tripartite représentant une folksonomie en deux sous graphes bipartites :

– Graphe bipartite tags- utilisateurs : ce graphe nous permet de regrouper les tags den classes en fonction des relations entre les utilisateurs.
– Graphe bipartite tags- ressources : Ce graphe permet de regrouper les tags en classes en fonction des relations d'utilisation de ces tags dans l'indexation de l'ensemble de ressources.

Nous déterminons ces clusters tout en groupant les tags de nature similaire ou ayant des caractéristiques communes ensemble. pour générer ces clusters, nous faisons appel aux techniques de détection de communautés où chaque communauté représente une classe de tags. Nous obtenons à la fin deux graphes purement composés que par des tags sur lesquelles nous appliquons l'algorithme de clustering. Pour faire des recommandations et pour déterminer les centres d'intérêts utilisateur, nous intéressons à déterminer les corrélations entre les clusters et les tags. La détermination des itemsets les plus fréquents, se fait par l'utilisation de l'algorithme Apriori Agrawal et al. (1993). L'objectif est de calculer les différentes corrélations entre les éléments (clusters et tags) en partant du plus général (cluster) au plus détaillé (tag). Nous nous basons sur le principe que si un cluster est fréquente alors elle contient forcément des tags fréquents. Le résultat de ce processus est l'ensemble de clusters les plus fréquentes et les plus corrélées entre elles. Par la suite nous passons à déterminer les règles d'association entre ces clusters résultats. Finalement, nous obtenons l'ensemble de règles d'associations entre les facettes les plus fréquentes. A partir de ces clusters corrélées, nous calculons les tags les plus corrélés de la même manière en appliquant l'algorithme Apriori et déterminer les règles d'associations ou de recommandation. L'utilisateur choisit le cluster ⌋ qui lui intéresse, notre système de recommandation interroge la base des règles pour vérifier sŠil existe une règle contenant comme partie prémisse le cluster ⌋. Quand une règle existe, il suggère à l'utilisateur les clusters représentés dans la partie conséquence de la règle sous la forme de nuage de tags présentant les principaux tags appartenant à chaque cluster. Au fur et au mesure de choix de l'utilisateur, la recommandation de deuxième niveau se déclenche de la même façon que celle de clusters.

4 Conclusion

Dans cet article, nous proposons une nouvelle approche de recommandation de tags. Notre modèle permet d'apprendre au cours du temps les habitudes et les préférences utilisateurs par l'exploitation de l'historique d'indexation. Cette approche permet de faciliter la tâche d'indexation de ressources en suggérant à l'utilisateur des recommandations qui répondent à ses besoins et ses attentes. Les perspectives de ce travail visent à améliorer la création de clusters en utilisant la technique de co-clustering selon plusieurs points de vues au même temps.

5 Summary

In this paper, we propose a new model of recommendation. Our approach is based on the structuring of tags already used by users to annotate their resources earlier in clusters. The recommendation comes at two levels : the recommendation and the recommendation of clusters of tags. To make recommendations more relevant, our approach involves the user to validate the recommendations of clusters initially proposed by the system. Once the user has made his choice of clusters, the system provides a list of rem mended of tags.

Références

Agrawal, R., T. Imieliński, et A. Swami (1993). Mining association rules between sets of items in large databases. In *SIGMOD '93 : Proceedings of the 1993 ACM SIGMOD international conference on Management of data*, Volume 22, New York, NY, USA, pp. 207–216. ACM.

Jäschke, R., L. Marinho, A. Hotho, L. Schmidt-Thieme, et G. Stumme (2007). Tag recommendations in folksonomies. In J. Kok, J. Koronacki, R. Lopez de Mantaras, S. Matwin, D. Mladenic, et A. Skowron (Eds.), *Knowledge Discovery in Databases : PKDD 2007*, Volume 4702 of *Lecture Notes in Computer Science*, pp. 506–514. Berlin / Heidelberg : Springer.

Jäschke, R., L. Marinho, A. Hotho, L. Schmidt-Thieme, et G. Stumme (2008). Tag recommendations in social bookmarking systems. *AI Communications 21*(4), 231–247.

Lu, Y.-T., S.-I. Yu, T.-C. Chang, et J. Y. jen Hsu (2009). A content-based method to enhance tag recommendation. In C. Boutilier (Ed.), *IJCAI*, pp. 2064–2069.

Pujari, M. et R. Kanawati (2012). Tag recommendation by link prediction based on supervised machine learning. In J. G. Breslin, N. B. Ellison, J. G. Shanahan, et Z. Tufekci (Eds.), *ICWSM*. The AAAI Press.

Xu, Z., Y. Fu, J. Mao, et D. Su (2006). Towards the semantic web : Collaborative tag suggestions. In *Proceedings of the Collaborative Web Tagging Workshop at the WWW 2006*, Edinburgh, Scotland.