



Open Archive TOULOUSE Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : <http://oatao.univ-toulouse.fr/>
Eprints ID : 13156

Video : <http://bda2013.univ-nantes.fr/index71da.html?q=node/24>

To cite this version : Mezghani, Manel and Amous, Ikram and Sèdes, Florence *An approach for social interest detection*. (2013) In: 29èmes Journées Bases de Données Avancées (BDA) Session Jeunes Chercheurs, 22 October 2013 - 25 October 2013 (Nantes, France). (Unpublished)

Any correspondance concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

An approach for social interest detection

Manel Mezghani^{1,2}

Supervisors: Ikram Amous¹ and Florence Sèdes²

Start date: December 2011 / **Funding source:** contract of higher education

¹ Sfax University, MIRACL Laboratory, Sfax, Tunisia

² Paul Sabatier University, IRIT Laboratory, Toulouse, France

mezghani.manel@gmail.com, ikram.amous@isecs.rnu.tn, Sedes@irit.fr

Abstract. In this article, we propose a new technique of interests detection by analyzing the accuracy of the tagging behaviour of each user in order to figure out the tags which reflect actually content of the resources. Our approach has been tested and evaluated in the Delicious social database.

Keywords: User interests, tagging behaviour, resources, social network.

I. Introduction

Social information is permanently growing. Consequently, adaptation systems using a user profile that reflects the appropriate characteristics (interests, preferences, etc.) could avoid cognitive overload and disorientation of the user when accessing the information space. In our work, we are interested in detecting the user's interests. This could be used in further works for an adaptation purpose.

Detecting user interests is a non trivial problem (Milicevic *et al.*, 2010). In fact, the user's profile building process suffers from the lack of information provided by himself. Indeed, the user generally doesn't give all the information related to his interests. So, his profile can never be considered fully known by a system. In order to overcome such a problem, the researchers have analyzed the social environment of the user such as his neighbours, his tagging behaviour, or even the objects (the resources) he interacts with.

According to (Astrain *et al.*, 2010), interests could be deduced from the social environment based on the user, the object or even the tag.

For the **user**, interests could be explicitly provided in the user's profile (Zayani *et al.* 2007), or implicitly deduced from his behaviour of navigation (Rebai *et al.*, 2012) or behaviour of tagging (Kim *et al.*, 2011). The user-based interest could be deduced from other users in the networks (neighbours). The neighbours could be deduced in different ways (Tchuente, 2013) (Kim *et al.*, 2011).

For the **object**, interests are deduced based on the objects that the user accesses (Ma *et al.*, 2011). To analyze resource content, different techniques exist such as the indexation which is used in order to extract the significant terms from resources. After indexing resources different scoring function could be applied in order to detect the most relevant resource according to a specific query (Vallet *et al.*, 2010).

For the **tag**, its utility has been proved to detect user's interest (Kim *et al.*, 2011). Tag-based interest detection could be deduced by analyzing used tags (De Meo *et al.* 2010) or by analyzing the semantic of tags (Kim *et al.*, 2011). Tag-based user profile modelling (interests, etc.) is detailed in (Mezghani *et al.*, 2012).

In this paper, we firstly show the differences of our approach compared to the other approaches in the same context. We then describe our proposal for detecting interests and the experiments

done to validate it. Finally, we conclude and discuss some future works.

II. Synthesis

After presenting some researches done to analyze the tagging behaviour elements, we now discuss the main differences between our approach and the other researches: i) Unlike most of researches which focus on the tag content considered as an interest (by analyzing the semantic of the tags for example), we will focus on analyzing the accuracy of the tags with regards to the resources content. ii) We focus on analyzing the object-based rather than the user-based interest detection. In fact object-based interest detection provides richer information than the user-based method (Song *et al.*, 2011). iii) For object-based interest detection, most researches do not consider the accuracy of the tags with the object (resource) content. This problem has been addressed in (Milicevic *et al.*, 2010). However, the proposed methods uses techniques such as clustering, semantic processing, etc. and none of them use the resources content analyze in their works.

To summarize, our approach uses the users' tags and treats them according to the content of their respective resources. The accurate tags are those reflecting the resources content. In order to validate our research, we will use the social environment that reflects the user's interests (Tchuente, 2013).

III. Proposed approach

In this section, we will propose our approach for detecting accuracy of the tags assigned by the users. These resulting accurate tags will be considered as relevant interests for the user. This approach is based on the hypothesis that a user, who tags a resource with keywords reflecting its content, is really interested in the thematic of this resource. This observation will be experimented and validated on the Delicious social dataset.

a. Description

In our approach, we analyze the tags assigned to the resources to detect user's interest. The resources are generally a set of URLs describing them. This process of detecting interest is applied to each user's neighbour. We extract in the first step the tagging behaviour relations, which is composed by the tags applied to the resources by each user.

In the second step, we extract the content of these URLs or web pages and index them as semi-structured files, using the

Lucene API¹. We will use it in order to figure out the most accurate tags with regards to the content of the tagged resource. Lucene relies on a field-based indexation technique. This characteristic enables indexing the documents according to one or more fields. For example, fields could be the *title*, the *content*, the *URL*, etc. In our approach, the indexation process has been made according to the *content* of the document. The indexing process is explained as follow: when Lucene indexes documents, it divides them into a number of terms. Then, it stores the terms in an index file, where each term is associated with the document contents. Terms are generated using an analyzer that converts each word in its root form. When a request is made it is treated by the same analyzer used to build the index and then used to find the corresponding term(s) in the index. This provides a list of documents matching the query.

After indexing the content of the resources, we assign a rank to each resource according to the assigned tag. This rank is computed from a function of similarity between the resource (as a XML file) and the query (as a tag). Many similarity functions exist in the literature such as the similarity function supported by Lucene². After ranking the resources, we test if the resource tagged by the query exists in the top-k result provided by the ranking function. If it's the case, we state the tag as relevant to the resource. This step is iterated for all tags of each user's neighbour. In order to validate that the relevant tags list are really pertinent for the user, we compare the founded relevant tags with the user's tag (real tagging behaviour).

b. Validation

We validate our approach upon the Delicious database that contains social networking, bookmarking, and tagging information. We have tested our approach on a set of 50 users. These users have different number of neighbours. The neighbours are the explicit friendship relation (the user's egocentric network). The method of validation uses the social environment of the user (the neighbours) to detect accurate interests. In fact the neighbours provide an information which reflect the user's interests (Tchuente, 2013). To validate that the tags list built from our approach can be accurate to a user, we compare the founded relevant tags of each user's neighbours with the tags in the user profile. A tag is stated accurate if it appears in the user's profile. The precision is calculated according to the number of accurate tags (which exist in the user's neighbours profile) divided by the total number of tags provided.

$$\text{Precision} = \frac{\text{Number of accurate results}}{\text{Number of accurate results} + \text{Number of inaccurate results}}$$

This precision is calculated for each single user's neighbours. The overall precision is the average of all the neighbours' precision. Figure 1 shows the overall precision, for this set of users, between the calculated relevant tags of his egocentric

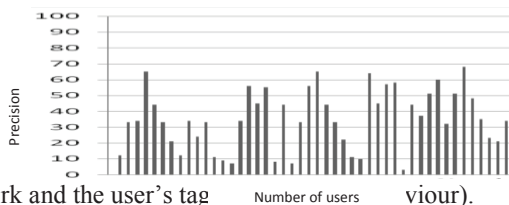


Figure 1: Precision according to the egocentric network of the accurate interests detected for a set of 50 users.

From this set of users, we have found that the precision varies according to different cases: i) for users who have a lot of friends, the precision is higher than those who have less friends. ii) for users who have a few friends, the test has provide a precision for a few users equal to zero. This means that the user doesn't share with his friend common interests. So, the users with a little number of friend relationships may influence negatively the accuracy of the interests detected.

The precision of the results demonstrate that our hypothesis is more correct when the users have more friends.

IV. Conclusion

In this paper, we have proposed a new approach of detecting social interests. This approach is based on analyzing the tagging behaviour of each user. In fact, the analysis aims to extract the most accurate tags according to their relevance to the content of the tagged resources. These latter, are indexed in order to extract meaningful terms. A score is applied to each tag according to its pertinence to the resources content. The relevant tags are those having a higher score. We have tested our approach on a set of users. We have also validated our approach by comparing the tags in the user profile with the tags stated as relevant of his explicitly friendship relation. In future works, we will test our approach on more users in order to have more scalable results.

References

- Astrain, J. J., Cordoba, A., Echarte, F. and Villadangos J. (2010): 'An algorithm for the improvement of tag-based social interest discovery'. SEMAPRO: The Fourth International Conference on Advances in Semantic Processing. 2010, pp. 49-54.
- De Meo, P., Quattrone, G. and Ursino D. (2010): 'A query expansion and user profile enrichment approach to improve the performance of recommender systems operating on a folksonomy'. User Modeling and User-Adapted Interaction. 2010, 20(1), pp. 41-86.
- Kim, H.-N., Alkhaldi, A., Saddik, A. E. and Joi G.-S. (2011): 'Collaborative user modeling with user-generated tags for social recommender systems'. Expert Systems with Applications, 2011, pp. 8488-8496.
- Ma, Y., Zeng, Y., Ren, X. and Zhong, N. (2011): 'User Interests Modeling Based on Multi-source Personal Information Fusion and Semantic Reasoning'. In Active Media Technology (AMT) 2011, pp. 195-205
- Mezghani, M., Zayani, C. A., Amous, I. and Gargouri, F. (2012): "A user profile modelling using social annotations: a survey". WWW (Companion Volume), 2012, pp. 969-976.
- Milicevic, A. K., Nanopoulos, A. and Ivanovic, M. (2010): 'Social tagging in recommender systems: a survey of the state-of-the-art and possible extensions'. Artif. Intell. Rev. Vol. 33 no. 3, 2010, pp. 187-209.
- Rebai Z, R., Zayani, C. A. and Amous, I. (2012): 'An Adaptive Navigation Method for Semi-structured Data'. Advances in Databases and Information Systems ADBIS (2), 2012, pp. 207-215.
- Song, Y., Zhang, L., and Giles, C. L. (2011): 'Automatic tag recommendation algorithms for social recommender systems'. ACM Trans., Web. Vol.5, no.1, Article 4, 2011, pp. 1-31.
- Tchuente, D., (2013): 'Modélisation et dérivation de profils utilisateurs a partir de réseaux sociaux : approche a partir de communautés de réseaux k-egocentriques'. Doctoral thesis, University of Toulouse, 2013.
- Vallet, D., Cantador, I. and Jose, J. (2010): 'Personalizing Web Search with Folksonomy-Based User and Document Profiles Advances in Information Retrieval'. Advances in Information Retrieval, 2010, Vol. 5993, pp. 420-431.
- Zayani, C. A., Péninou, A., Marie-Françoise, C. and Sedes, F. (2007): 'Towards an Adaptation of Semi-structured Document Querying'. Proceedings of the CIR'07 Workshop on Context-Based Information Retrieval CIR 2007, CEUR-WS.org, Vol. 326.

¹ <http://lucene.apache.org/>

² <http://ipl.cs.aueb.gr/stougiannis/default.html>