IJCSS International Journal of Complex Systems in Science

Int. J. Complex Systems in Science
vol. **1(2)** (2011), pp. 134–140

Integration of semantic user profile within a social recommendation system: semantic relevance measures characterization

Maria Malek[1,†], Dalia Sulieman[1,2] and Hubert Kadima[1]

[1] *LARIS-EISTI , Ecole Internationale de Sciences de Traitement de l'Information Cergy-Pontoise France*

[2] *ETIS, University of Cergy Pontoise France*

Abstract. We have earlier proposed an algorithm for computing recommendation in collaboration networks. This recommendation is based, on one hand, on the similarity between authors' preferences and the submitted query and on the other hand, on the betweenness centrality of authors found on the search paths.

We are now working on the elaboration of semantic user preferences by using a domain ontology.

We propose different categories of relevance measures. The basic relevance measure is the classical one which matches the preference terms to the query terms. The semantic relevance measure is the one that matches the semantic preferences to the query concept terms both extracted from the ontology definition. The semantic-structural relevance measure integrates the structural and hierarchical organization of concepts in the ontology.

† **Corresponding author:** maria.malek@eisti.fr
Received: October 14, 2011
Published: October 24, 2011

1. Introduction

Social networks can be a source for the development of recommendations like finding an expert in a given field, suggesting products to sell, offer a friend, etc. This development may be based on paths exploration algorithm and degree analysis algorithms[9, 1, 2, 3, 10, 5].

In this paper, we propose an algorithm for computing recommendation in collaboration networks. Our network consists of a set of persons with weighted ties that represent similarity between actors. We work on two types of collaboration networks:

1. A bibliographic collaboration networks which is composed of authors related together by similarity links. These authors are extracted from the Microsoft bibliographic data. The similarity between two authors depends on the co-citation and the bibliographic coupling measures [7].
2. A product co-purchasing collaboration network extracted from the Amazon data set. This data set contains 548551 products.

To answer the request of an actor, the system recommends a list of other actors that match the best requested criteria.

We propose two recommendation algorithms based on three types of knowledge:

- The first type deals with information concerning the person. This information is stored in the actor vertex level and can be represented by an ontology describing user profiles.
- The second type of information is computed from the network structure itself. Actually, this consists of exploring the links starting from the initial actor exploring the maximum spanning tree whose the root is the initial actor. We can thus reduce the search space of target actors.
- The third type of information is based on the betweenness centrality measure associated to each actor. This measure enables to estimate the control of an actor over other pairs of actors. We use this measure to extract the best paths from the previous spanning tree.

we have earlier proposed an algorithm for computing recommendation in collaboration networks [4, 6]. This recommendation is based, on one hand, on the similarity between authors' preferences and the submitted query and on the other hand, on the betweenness centrality of authors found on the search paths. To search the graph we first extract the most representative spanning tree and then we explored this tree.

The first proposed algorithm is an exhaustive one, it is based on breadth strategy to explore the spanning tree until finding a suitable author to recommend. The second algorithm uses the A* algorithm for searching the spanning tree instead of the breadth search strategy. By comparing both algorithms we notice that 11% to 49% of the original search space is explored.

2. Extension of user preferences by using a domain ontology

We use a domain ontology which is a taxonomy as in the Amazon application. The single relation between some pairs of concepts is the relation "Is-a". Such an ontology is defined by a set of concepts terms: $\{C_1, C_2, .., C_n\}$ and a set of pairs of concepts related by the relation Is-a. This ontology can be represented by a tree and forms a taxonomy. We aim to integrate the ontology concepts terms as well as its structure in order to enhance the recommendation quality [8]. As mentioned above, the user preferences are actually represented by a weighted vector of terms. We propose to extend the user preferences by annotating this weighted vector of terms, using the ontology concept terms; we thus can derive the semantic profile of user which will be represented by a weighted vector of concepts. For example, suppose that the user u_1 has as basic preferences the weighted vector: $\{(I_1, S_1), (I_2, S_2), (I_3, S_3), (I_4, S_4)\}$ where I_i is a given item and S_i is the score associated to this item. We suppose that these items belong to concept terms of the ontology domain : $\{(I_1, C_2), (I_2, C_3), (I_2, C_4), (I_3, C_4), (I_4, C_2)\}$. The user semantic derived preferences will be : $\{(C_2, \frac{S_1+S_4}{2}), (C_3, S_2), (4_4, \frac{S_2+S_3}{2})\}$.

We therefore assign to each user two types of preferences:

1. basic preferences represented by a weighted list of terms named: BPref,
2. semantic preferences represented by a weighted list of concepts named CPref.

Our goal is to elaborate set of a similarity measures in order enrich the semantic part of our algorithm and to test the improvement of the recommendation quality.

We propose theses categories of relevance measures:

1. The basic relevance measure: this is the measure that we use actually, the idea is to compute how match the query is close to the a *basic user preferences*. Given a query R_X composed of a list of terms, the basic

relevance is computed by:

$$rel(R_X, BPref_Z) = \frac{\sum\limits_{j \in inter(R_X, BPref_Z)} BPref_Z.S_j}{\sum\limits_{i=1}^{m} BPref_Z * S_j + |R_X \backslash BPref_Z|} \tag{1}$$

With:

$$inter(R_X, BPref_Z) = \{k \in \{1, \dots m\} \mid BPref_Z \cdot T_K \in R_X\}$$

In the numerator part of the equation, we compute a value that depends on the weighs of terms that belongs to the query and to the user preferences at the same time. In the denominator part, the union of terms of the query and the user preferences is computed.

2. The semantic relevance measure: we propose a simple extension of the basic similarity measure this measure between a semantic representation of the query and the semantic user preferences profile. Given a query by a list of terms we derive a semantic representation of a query by: $CR = \{C_1, ..C_m\}$, where $C_1, .., C_m$ are the concepts tow which belong the terms of query. The similarity measure computed by:

$$rel(R, CPref_Z) = \frac{\sum\limits_{j \in inter(CR, CPref_Z)} CPref_Z.S_j}{\sum\limits_{i=1}^{m} CPref_Z * S_j + |CR \backslash CPref_Z|} \tag{2}$$

With:

$$inter(CR, CPref_Z) = \{k \in \{1, \dots m\} \mid CPref_Z \cdot T_K \in CR\}$$

3. The semantic-structural relevance measure: this type of measure integrates the ontology structure beside the concepts definitions. An example of this measure is the distance proposed by [8] between two concepts x and y that belong to a given taxonomy (a domain ontology as the Amazon one).

$$D(x, y) = \frac{log(1 + 2\beta(y, LCA(x, y))) - log(\alpha(x, LCA(x, y)))}{maxD} \tag{3}$$

where :

- $LCA(x, y)$ is the least common ancestor between x and y.

- $\alpha(x, y)$ is a generalization measure that describes the common characteristics between x and y.
- $\beta(y, x)$ is a specialization measure that describes the characteristics of y but not of x.

These two last measures are given by:

$$\alpha(X, Y) = \frac{APS(Y)}{APS(X)}$$

$$\beta(Y, X) = APS(Y) - APS(X)$$

where APS(X) means the a-priori score of concept node and is given by:

$$APS(X) = \frac{1}{n_c + 2}$$

and n_c is the number of descendants of X.

3. Conclusion and perspective

we have earlier proposed an algorithm for computing recommendation in collaboration networks [4, 6]. This recommendation is based, on one hand, on the similarity between authors' preferences and the submitted query and on the other hand, on the betweenness centrality of authors found on the search paths. To search the graph we first extract the most representative spanning tree and then we explored this tree.

The first proposed algorithm is an exhaustive one, it is based on breadth strategy to explore the spanning tree until finding a suitable author to recommend. The second algorithm uses the A* algorithm for searching the spanning tree instead of the breadth search strategy.. By comparing both algorithms we notice that 11% to 49% of the original search space is explored.

We are now working on the elaboration of semantic user preferences by using a domain ontology. The Amazon data sets (http://snap.stanford.edu/) contains 548551products described by:

- Two identifiers: Id , ASIN: Amazon Standard Identification Number.
- Title,group: (Book, DVD, Video or Music), salesrank.
- Similar: ASINs of co-purchased products.
- Categories: location in product category hierarchy.

- Reviews: Product review information: time, user id, rating, total number of votes on the review, total number of helpfulness votes (how many people found the review to be helpful).

The data preparation process consists of:

1. Elaboration of the products taxonomy.

2. Extraction of the collaboration network (nodes are users).

3. Elaboration of the basic and semantic preferences for users.

We propose different categories of relevance measures. The basic relevance measure is the classical one which matches the preference terms to the query terms. The semantic relevance measure is the one that matches the semantic preferences to the query concept terms both extracted from the ontology definition. The semantic-structural relevance measure integrates the structural and hierarchical organization of concepts in the ontology. We aim now to test the different types of relevance measures in order to propose a metric that enables to elaborate the best recommendation.

References

[1] Lada A. Adamic, Orkut Buyukkokten, and Eytan Adar. A social network caught in the web. *First Monday*, 8(6), 2003.

[2] Christopher S. Campbell, Paul P. Maglio, Alex Cozzi, and Byron Dom. Expertise identification using email communications. In *CIKM*, pages 528–531, 2003.

[3] Yupeng Fu, Rongjing Xiang, Yiqun Liu, Min Zhang, and Shaoping Ma. Finding experts using social network analysis. In *Web Intelligence*, pages 77–80, 2007.

[4] Hubert Kadima and Maria Malek. Toward ontology-based personalization of a recommender system in social network. In *International Conference on Soft Computing and PAttern Recognition (SoCPaR 2010), IEEE*, December 2010.

[5] Yao Lu, Xiaojun Quan, Xingliang Ni, Wenyin Liu, and Yinlong Xu. Latent link analysis for expert finding in user-interactive question answsering services. In *2009 fifth International Conference on Semantics, Knowledge and Grid*, 2009.

[6] Maria Malek and Dalia Sulieman. Exhaustive and guided algorithms for recommendation in a professional social network. In *7th conference on Application of Social Network Analysis (ASNA), Zurich*, Sepetember, 2010.

[7] M. E. J. Newman. Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Science of the United States (PNAS)*, 101:5200–5205, 2004.

[8] Vincent Schickel-Zuber and Boi Faltings. Oss: A semantic similarity function based on hierarchical ontologies. In *IJCAI*, pages 551–556, 2007.

[9] Jun Zhang and Mark S. Ackerman. Searching for expertise in social networks: a simulation of potential strategies. In *GROUP*, pages 71–80, 2005.

[10] Jun Zhang, Mark S. Ackerman, and Lada Adamic. Expertise networks in online communities: structure and algorithms. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pages 221–230, New York, NY, USA, 2007. ACM.