Djuana Tjhwa, Endang, Xu, Yue, Li, Yuefeng, & Josang, Audun
(2011)
Ontology learning from user tagging for tag recommendation making.
In Petit, J M, Hubner, J F, & Suzuki, E (Eds.) *Proceedings of the 2011 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Workshops.*
Institute of Electrical and Electronics Engineers Inc., United States, pp. 310-313.

https://doi.org/10.1109/WI-IAT.2011.163

Learning Domain Ontology for Tag Recommendation

Endang Djuana, Yue Xu, Yuefeng Li

ABSTRACT

Recently, user tagging systems have grown in popularity on the web. The tagging process is quite simple for ordinary users, which contributes to its popularity. However, free vocabulary has lack of standardization and semantic ambiguity. It is possible to capture the semantics from user tagging and represent those in a form of ontology, but the application of the learned ontology for recommendation making has not been that flourishing. In this paper we discuss our approach to learn domain ontology from user tagging information and apply the extracted tag ontology in a pilot tag recommendation experiment. The initial result shows that by using the tag ontology to re-rank the recommended tags, the accuracy of the tag recommendation can be improved.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Information filtering

General Terms

Algorithms, Experimentation

Keywords

User tagging, ontology learning, tag recommendation

1. INTRODUCTION

User tagging or collaborative tagging describes the process by which many users add metadata in the form of keywords to Internet resources with a freely chosen set of keywords (tags) [2].

The tagging process is quite simple for ordinary users who do not need to have systematic classification background which brought to its popularity. However, free and relatively uncontrolled vocabulary has its drawback in terms of lack of standardization and semantic ambiguity. Three of these problems are polysemy, synonymy, and basic level variation [2]. Also, the flat and non-hierarchical structure leads to low search precision and poor resource navigation.

Collaborative tagging systems usually include tag recommendation mechanism to assist with the process of finding good tags for an item. To be able to recommend the most relevant tag, the semantic meaning of tags used by users and especially the **semantic relationships** between tags in the tag collection should be taken into consideration. So far, semantic relationships between tags have not been sufficiently exploited in the existing works. These problems motivate the work we introduce in this paper that aims to represent the semantic meaning and relationship of tags for the purpose of making recommendation.

In this paper we present our approach to ontology learning from user tagging information and its application for improving tag recommendation in a pilot experiment. We begin by providing a bit of backgrounds in Section 2. We then introduce our ontology learning approach in Section 3. In Section 4 we discuss the proposed improvement and initial results. In Section 5 we review related works. Section 6 concludes this paper and gives some ideas for further work.

2. BACKGROUNDS

2.1 User Tagging Collection

User tagging collection consists of three entities which are items (or resources), tags assigned to these items and users who assign these tags to the items. Tags are typically arbitrary strings (which could be a single word or short phrase). In this paper, a tag is defined as a sequence of terms where a term is any word.

Based on the three entities of user tagging collection, the collaborative tagging system is formulated as 4-tuple: $F \coloneqq (U, T, I, Y)$ by Jaschke et al. [3] where: U, T, I are finite sets, whose elements are the users, tags and items respectively, and Y is a ternary relation between them, i.e., $Y \subseteq U \times T \times I$, whose elements are called tag assignments (or tas for short). An element $(u, t, i) \in Y$, represents that user u collected item i using tag t. A function $F(u, i)$ is defined to return a set of tags that a user u has assigned to an item i: $F(u, i) = \{t \in T | (u, t, i) \in Y\}$ for all $u \in U$ and $i \in I$.

2.2 Tag Recommendation

A tag recommender is a specific kind of recommender systems in which the goal is to suggest a set of tags to use for a particular item to a user during the annotation process. Based on previous formulation of collaborative tagging system the task of a tag recommender system is to recommend, for a given user $u \in U$ and a given item $i \in I$ with $F(u, i) = \emptyset$, a set $\tilde{T}(u, i) \subseteq T$ of tags. In many cases $\tilde{T}(u, i)$ is computed by first generating a ranking on the set of tags according to some quality or relevance criterion from which then the top n tags are selected [3].

3. ONTOLOGY LEARNING

In this work we propose to construct the tag ontology based on backbone ontology. We mapped the tags in the tag collection to the concepts on the backbone ontology and make use of the available relationships among concepts in the backbone ontology. We chose WordNet foundational ontology [1] as the backbone ontology as it has wide coverage of concepts (over 200,000) and richness of relationships as well as availability of accompanying corpus and other facility for disambiguation process.

Two main tasks are included in the proposed tag ontology construction: to find the meaning of user tags and to find the relationships among tags. For the first stage, disambiguation is needed to identify the most relevant concept for a tag. The second stage involves finding all the relation between the mapped concepts by going through the hierarchy in the backbone ontology for semantic relationships such as “is-a” or “part-of”.

3.1 Ontology Definition

The backbone ontology is formally defined as a 2-tuple $BackboneONTO := (C, R)$ where C is a set of concepts; R is a set of relations representing the relationships between concepts.

A concept c in C is a 2-tuple $c = (id, synset)$ where id is a unique identification assigned by WordNet to the concept c; and $synset$ is a synonym set containing synonymic terms which represent the meaning of the concept c. For easy to describe the work, we denote the the identifier of a concept c by $id(c)$; and the set of synonyms representing c by $synset(c)$.

Let $S = \{w | \exists c \in C, w \in synset(c)\}$ be the set of all synonymic terms. For the terms in a $synset$, each term $w \in synset(c)$ is a 2-tuple $(w, freq_c(w))$ where w is a synonym in the $synset$; $freq_c(w)$ is the frequency assigned by WordNet to the term as an indication of how frequently this term has been used to represent the meaning of the concept based on the accompanying corpus.

A relation r in the relation set R is a 3-tuple $r = (type, x, y)$, where $type \in \{is\_A, part\_Of\}$; x, y are the concepts that hold the relation r.

3.2 Mapping tags and Disambiguation

One tag may contain one or more terms. It is possible that a tag can be mapped directly to one or more concepts in the backbone ontology. It is also possible that only part of a tag may map to one or more concepts. We propose the following mappings to deal with different cases:

1. Direct mapping: For each tag, we try to map the tag as a whole to the concepts in the backbone ontology. If the tag is a synset term of a concept, the concept is considered a mapping of the tag. We define the following function to represent the mapping from a tag to concepts: $Tag\_Concept: T \rightarrow 2^C$. $\forall t \in T,\ Tag\_Concept(t) = \{c | c \in C, \exists (w, f) \in synset(c), t == w\}$ is a set of concepts for each of which t is one of its synset terms.
2. Partial Mapping: When a tag could not be directly mapped we firstly conducted phrase shortening by one word at a time from start of phrase to the end to see if in any stage we can map the shortened phrase.
3. Term Mapping: For each of the remaining tags, we conducted the split tag mapping. We first map each of the terms to a concept, then we conduct a disambiguation process.

After all the possible mapped concepts are found for a tag, we need to choose the most appropriate concept from the mapped concepts to represent the meaning of the tag. We define a Term_Concept matrix: $T\_C[t_i, c_j]_{m \times n}$ to represent the strength of the mapping between tags and concepts, where m=|T| and n=|C|. The initial matrix is generated during the mapping process and the initial mapping strength is the word frequency associated with the term. In order to make the frequency comparable between different concepts and terms, we normalize the frequency value to a scale of [0, 1]. $T\_C[t_i, c_j]$ is modified to provide the normalized frequency instead of the original term frequency:

$$T\_C[t_i, c_j] = \begin{cases} \dfrac{freq_{c_j}(t_i)}{\sum_{c_k \in Tag\_Concept(t_i)} freq_{c_k}(t_i)} & c_j \in Tag\_Concept(t_i) \\ 0 & otherwise \end{cases}$$

4. PROPOSED IMPROVEMENT

In order to evaluate the potential improvement, we implemented a baseline tag recommender system proposed in [3] which is based on the user-based collaborative filtering (CF) method. The recommendation result of this baseline system is then modified based on the tag relation information obtained from the tag ontology. The original recommendations and the modified recommendations are compared to indicate the improvement achieved by using the proposed tag ontology.

In the baseline tag recommender the ranking calculation conducted may result in a tie. Ties between ranking values in most cases were solved by random selection. This leads to uncertainty of ranking which can leads to a good tag being missed out due to random selection process. Based on this we proposed a possible improvement to the potential ranking tie problem based on a re-ranking approach according to semantic relations in the extracted ontology.

The re-ranking approach compares the relative distance between the recommended tags to determine if one tag is more specific or more general in the ontology hierarchy. We assign a score based on this relative position to each tag. The more specific one tag, the higher the score is to this tag, and the higher the rank is. The initial experiment result shows that the recommendation results can be improved.

5. RELATED WORKS

There are several works which tried to extract ontological structures from user tagging systems. Lin et al [4] extracted ontological structures by exploiting low support association rule mining. Trabelsi et al [5] focused more on extracting non-taxonomic relationships from folksonomies using triadic concepts.

6. CONCLUSION AND FURTHER WORK

We have discussed our approach to ontology learning from user tagging and presented the potential improvement to tag recommendation problem by improving the ranking of recommendations list. There is opportunity to improve the recommendation by exploiting further the extracted ontology structure for instance by considering the distance among concepts to find more neighbors and reducing the sparsity problem.

7. ACKNOWLEDGMENTS

This work is part of ARC Linkage Project (LP0776400) supported by the Australian Research Council.

8. REFERENCES

[1] Fellbaum, C. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA.

[2] Golder, S. and Huberman, B. 2006. *The structure of collaborative tagging systems*. Technical Report. HP Labs. http://www.hpl.hp.com/research/scl/papers/tags/tags.pdf.

[3] Jaschke, R., Marinho, L., Hotho, A., Schmidt-Thieme, L. and Stumme, G. 2008. Tag recommendations in social bookmarking systems. *AI Communications.* (2008), 231-247.

[4] Lin, H., Davis, J., and Zhou, Y. 2009. An integrated approach to extracting ontological structures from folksonomies. *The Semantic Web: Research and Applications.* (2009), 654-668.

[5] Trabelsi, C., Jrad, A. B. and Yahia, S. B. 2010. Bridging folksonomies and domain ontologies: Getting out non-taxonomic relations. In *Proceedings of IEEE International Conference on Data Mining Workshops.* (2010), 369-379.