Linking Images to Semantic Knowledge Base with User-generated Tags

Shuangyong Song
Information Technology Laboratory,
Fujitsu R&D Center Co., Ltd.
Beijing P.R. China 100027
shuangyong.song@cn.fujitsu.com

Qingliang Miao
Information Technology Laboratory,
Fujitsu R&D Center Co., Ltd.
Beijing P.R. China 100027
qingliang.miao@cn.fujitsu.com

Yao Meng
Information Technology Laboratory,
Fujitsu R&D Center Co., Ltd.
Beijing P.R. China 100027
mengyao@cn.fujitsu.com

ABSTRACT

Images account for an important part of Multimedia Linked Open Data, but currently most of the semantic relations between images and other entities are based on manual semantic annotation. With the popularity of image hosting websites, such as Flickr, plentiful tagging information of images makes it possible to automatically generate semantic relations between images and other semantic entities. In this paper, we propose a model for linking images to semantic knowledge base (*KB*) with user-generated tags of those images, while taking into account topical semantic similarity between tags. The experimental results show that our approach can effectively realize the mentioned aim.

CCS Concepts

• Knowledge representation and reasoning→Semantic networks.

Keywords

Semantic web; Multimedia *LOD*; Knowledge base; Social Tagging Systems; Web image; User-generated tags; Topic model; Social Network.

1. INTRODUCTION

Images constitute an important part of the Multimedia Linked Open Data (Multimedia *LOD*), but how to share and search images on the semantic web remains a significant yet challenging research issue [1]. Manual semantic annotation of images provides an opportunity to make the semantics of an image explicit and richer, and most of the semantic relations between images and other entities are based on manual semantic annotation [2]. Some researches try to automatically tag web images based on their web page text [3]; however, the complexity of web text generates much noise data and makes it difficult to detect the exact text that are really related to a target image.

With the development of Web2.0, online image tagging systems, such as Flickr, have attracted great attention in that they enable an effective way for users to organize and share images. Researches on understanding images based on their tags have been attached importance [4], so in this paper we aim to create links between images and semantic knowledge by using those tagging information of images, in this way to realize semantic image retrieval and detection of image-related relations.

In our proposal, we firstly detect the image tags which can be unambiguously linked to high-quality public semantic knowledge and process some ambiguous tags with simple tags' Co-occurrence relation, and then use the 'image-tag' relation matrix to train topic models, through which to calculate topical semantic relations between image tags and detect implicit semantic links between tags and semantic knowledge. Finally, we link images to semantic *KB* through tag-based mediation.

Yang et al. [5] exploited vocabulary-based knowledge when topically modelling short texts, so as to improve their readability. However, this measure is unavailable in our current task. Tags are quite different from words as a fraction of tags may be acceptable words, but a bulk of them often become non-verbal tags such as "mcm", "whatweshowtotourists" and "gsx1300r". Therefore, vocabulary-based methods are restricted on tag analysis because they may lead to different treatments on verbal and non-verbal tags. For these reasons, we exploit vocabulary-based knowledge on the step of tags' semantic link detection instead of the topic modelling part. In the following parts of this article, we will illustrate the details of the problem definition and proposed model, and report the experimental results.

2. PROBLEM DEFINITION & APPROACH

2.1 Problem Definition and Framework Description

Referring to the description in [2], we give the definition of an *image ontology* as 'an ontology of this image with a unique URI and available links to public recognized semantic *KB*', and assume that an image without available links to semantic *KB* should not be converted into an image ontology. We use *DBpedia* as referred semantic *KB*, and take *dbo* as acronym for *DBpedia* Ontologies <http://dbpedia.org/ontology/>, and take *dbr* as acronym for *DBpedia* Resources <http://dbpedia.org/resource/>. In *dbo*, ontologies include some category-level knowledge definitions, such as <http://dbpedia.org/ontology/movie> ('dbo:movie' for short) and 'dbo:movie', etc. In *dbr*, resources include some instance-level knowledge definitions, such as <http://dbpedia.org/resource/The_Shawshank_Redemption_(film)> ('dbr:The_Shawshank_Redemption_(film)' for short) and 'dbr:Somophyllin_CRT', etc. We define the set of images as $M = \{m_1, m_2, \ldots, m_i, \ldots, m_I\}$, and for each image m_i, a series of user-defined tags are given as $T(m_i)$. Given an image m_i and its related tags $T(m_i)$, our goal is to create an image ontology $O(m_i)$ by linking m_i to *dbo* or *dbr*, while considering both explicit and implicit semantic links between $T(m_i)$ and *dbo* or *dbr*.

SEMANTiCS'16, September 12–15, 2016, Leipzig, Germany.

DOI: http://dx.doi.org/10.1145/2993318.2993340

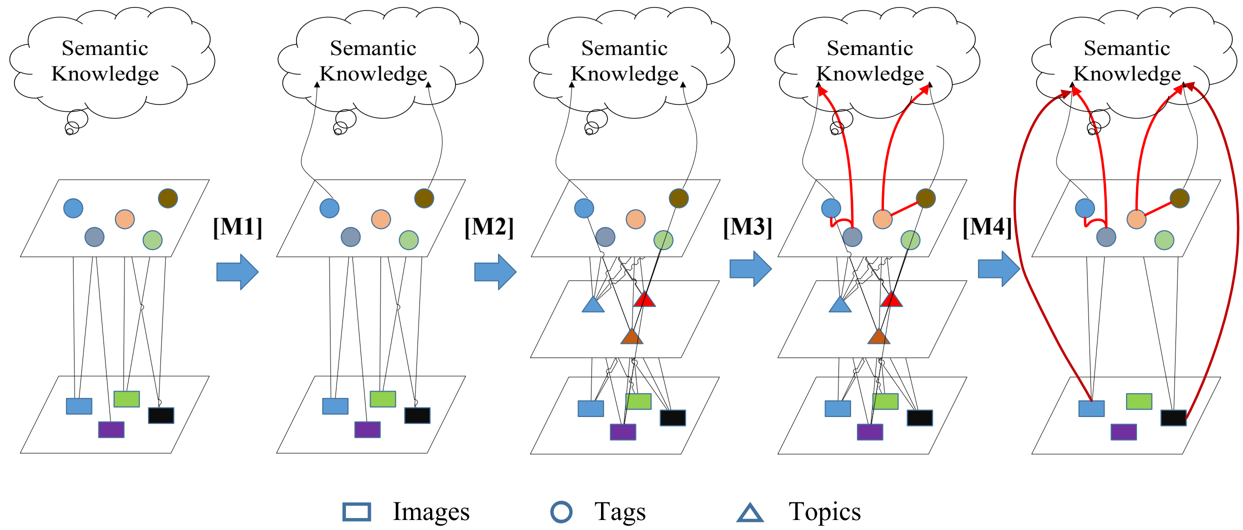

Figure 1. The framework of the proposed model.

Figure 1 shows the framework of our proposed model. It contains four functional modules: **[M1]** linking tags to knowledge; **[M2]** detecting semantic relations between tags; **[M3]** extending knowledge linked tags; **[M4]** linking images to knowledge. The main idea of our framework is creating links between images and semantic knowledge with taking image tags as effective medium. The mechanism of each functional module in our proposed framework is discussed in detail in the following subsections.

2.2 Linking tags to knowledge

In this step, we firstly link all unambiguous tags to *dbo* or *dbr*, which means that a tag has one and only one matched knowledge definition in *dbo* or *dbr*. For example, when $T(m_i)$ contains a tag 'Somophyllin-DF', we search for matched ontology with 'Somophyllin-DF' in *dbo* or *dbr*, and detect the unambiguous matched ontology as 'dbr:Somophyllin-DF', then we add it to $O(m_i)$. Likewise, those unambiguous matched tags include 'creditcard', 'CEO', 'governmentbuilding', and 'singer', etc. However, when an image is tagged with 'Apple', we may not be able to link it with 'dbr:Apple_Inc.' or 'dbr:Apple_III' or some other ontologies which related to 'Apple'. So when we face those ambiguous tags, we design a simple tags' Co-occurrence relation based method to determine linking an ambiguous tag to which ontology or not linking it to any ontology. The formula of Co-occurrence relation between an ambiguous tag t_i^a and an unambiguous tag t_j^u is given below:

$$R(t_i^a,t_j^u)=\sum_{k=1}^{U}R(t_i^a,t_k^u,t_j^u)=\sum_{k=1}^{U}\frac{C(t_i^a,t_k^u)*C(t_k^u,t_j^u)}{F(t_k^u)} \qquad (1)$$

where $R(t_i^a,t_j^u)$ means Co-occurrence relation between t_i^a and t_j^u, and $R(t_i^a,t_k^u,t_j^u)$ means partial Co-occurrence relation between t_i^a and t_j^u created through t_k^u. U means number of all unambiguous tags. $C(t_i^a,t_k^u)$ means Co-occurrence frequency of t_i^a and t_k^u, and $F(t_k^u)$ means frequency of t_k^u. Finally, a t_j^u with the maximum $R(t_i^a,t_j^u)$ with t_i^a with be detected as vicarious tag of t_i^a.

2.3 Detecting semantic relations between tags

Probabilistic topic models have been proved to be powerful tools for identifying latent text patterns in the content. Especially, Latent Dirichlet Allocation (*LDA*) [6] achieves the capacity of generalizing the topic distributions so that the model can be used to generate unseen documents as well. *LDA* was also applied and demonstrated usefulness in various work on semantic dimension reduction. Thus, we apply *LDA* on detecting topical information of image tags.

We utilize *JGibbLDA* version topic model to detect topical information from image-tag matrix, and this topic model is trained off line. Since the image-tag matrix is sparse and difficult to be well analyzed, topical information can help to discover the implicit semantic relations between tags. For example, it helps us to detect the relation between tags 'programmer' and 'ilife' since those two tags have very similar topical distribution vectors. Besides, this step can be regarded as a dimension reduction of tag vector space, which can greatly reduce computational consumption of topical semantic similarity between tags. If we set the number of topics as K, then each tag can be represented as a K-dimension vector.

2.4 Extending knowledge linked tags

The aim of this step is to expand the scope of 'knowledge linking tags' by considering topical semantic similarities between tags and some lexical analysis. Lexical analyses include synonyms analysis, plurals analysis and gerund analysis, while topical semantic similarity based method focuses on topical vectors of tags. For checking the possibility of linking an unlinked tag to semantic *KB*, we calculate its topical semantic similarity with all linked tags with cosine-similarity between their topical vectors [8], and choose those with similarity greater than threshold σ to be synonymous tags, where σ = 1-10^{-d}, and d is a positive integer, while a larger d means a stricter threshold. Besides, for two tags t_1 and t_2, supplemented with 'inclusion relation' and 'Levenshtein distance' between them, we design some rules for judging if they are similar tags. The rules are as below:

1) If t_1 and t_2 have 'inclusion relation', such as 'motor' and 'motorcycle', and cosine-similarity value between their topical vectors is bigger than σ, we judge them as similar tags;

2) If 'Levenshtein distance' value between t_1 and t_2 is equal to or smaller than a threshold β, which is a positive integer, and cosine-similarity value between their topical vectors is bigger than σ, we judge them as similar tags;

3) If t_1 and t_2 don't have above relations, we judge them as dissimilar tags.

If t_1 and t_2 are judged as similar tags, and just one of them has semantic link to knowledge base, we link the other one to the same knowledge with a probability, of which the value equals to $C(t_1,t_2)$, which is the cosine-similarity value between t_1 and t_2.

2.5 Linking images to knowledge

Based on mapping relations between tags and dbo or dbr, we link images to knowledge by taking tags as medium. In this step, the key part is selection of linking types, which is called predicate in <*subject*, *predicate*, *object*> triples. We use 77 million triples collected from our semantic database '*LOD4ALL*' [7] as criteria for predicate selection. We create links between images and objects with unambiguous predicate. For example, we use predicate 'dbo:locationCountry' to create links to objects such as 'China' or 'Denmark', and use predicate 'dbo:DigitalCamera' for links to objects such as 'Conon' or 'Sony'. For objects with multiple predicates, links will not be created while for other unknown objects, we tentatively use predicate as <http://www.w3.org/2000/01/rdf-schema#seeAlso>, which is shortened as 'rdfs:seeAlso' and means 'further information about the subject resource'.

In particular, we also detect some tag-combined knowledge for expanding the links' range. For example, if an image has both 'DigitalCamera' and 'Conon' as its tags, we will check if there is 'DigitalCamera_Conon' or 'Conon_ DigitalCamera' in *dbo* or *dbr*, and link this image to the detected knowledge.

3. EXPERIMENTS

3.1 Dataset and Parameter Settings

To evaluate the performance of the proposed model, a dataset from *mirFlickr* [4] was utilized, which contains 690,649 images and 659,227 tags. After removing 41,127 images without any tag and 392 attached with 'stopword' tags, we had 649,130 images left, and those images and their tags were used to train the topic model with an empirical K value of 200. Besides, all the 7,401 ontologies in *dbo* and 7,925,232 resources in *dbr* were used as semantic *KB*.

For evaluating the effects of different β, we set β as positive integer between 1 and 6 and compare the performance with different β, since we empirically judge that two tags with 'Levenshtein distance' more than 6 are with little probability to be similar tags. We randomly choose 500 tag-couple with 'Levenshtein distance' equals to or smaller than 6 as the dataset for evaluating different β, and we roughly set the recall as 1.0 when β = 6 considering we are unable to collect all valid tag-couples. We firstly get the *recall* and *precision* results and further the *F1*-value to evaluate β. Table 1 shows the results. As shown, we get the best performance when β is 2. Therefore, we set β = 2 in the following experiments.

Table 1. The results of different values of β for determination of tags' relation with 'Levenshtein distance'

	$\beta = 1$	$\boldsymbol{\beta = 2}$	$\beta = 3$	$\beta = 4$	$\beta = 5$	$\beta = 6$
Precision	0.907	0.900	0.692	0.438	0.289	0.240
Recall	0.567	0.600	0.692	0.767	0.842	1.000
***F1*-value**	0.697	**0.720**	0.692	0.558	0.428	0.387

Totally 124,095 tags are unambiguously mapped to *dbo* or *dbr*, which only account for 18.82% of all tags. To evaluate the effects of different d when β = 2, we manually label 200 images. We roughly set the recall as 1.0 when d = 1 considering we are unable to collect all possible related knowledge to an image. After getting both *precision* and *recall* with different d, we also use *F1*-value as evaluation criterion to evaluate d. Table 2 shows the results. As shown, the best performance shows when d = 3, which indicates that when $\sigma = 1\text{-}10^{-3}$ we can get best discriminant performance. Therefore, we set d = 3 in the following experiments.

Table 2. The results of different values of d for our model

	$d = 1$	$d = 2$	$\boldsymbol{d = 3}$	$d = 4$	$d = 5$	$d = 6$
Precision	0.317	0.531	0.933	0.889	1.000	1.000
Recall	1.000	0.895	0.737	0.421	0.158	0.053
***F1*-value**	0.481	0.667	**0.824**	0.571	0.273	0.100

3.2 Experimental Results

With our model, 535,622 tags are finally mapped to *dbo* or *dbr*, which account for 81.25% of all tags, which makes tremendous growth compared to 18.82%. In this subsection, to evaluate the performance of the proposed model on linking images to semantic knowledge base, we compare it with two other baseline models:

• Word2vec based model (*W2V*): Word2vec is a recently popular method for getting distributed representation of words, of which the output form is similar to *LDA*. However, we suppose that *W2V* is unfit for the image-knowledge linking task, since *W2V* does well in detecting context similarity, such as the similarity between 'Paris' and 'Beijing'. We will evaluate it in the following part.

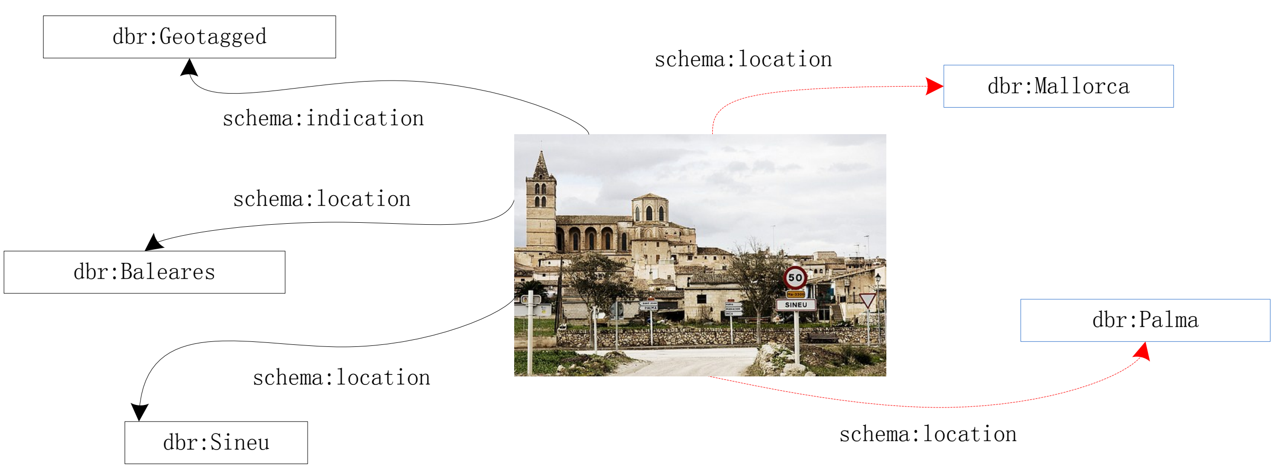

Figure 2. Example of a semantic image ontology (partial) about buildings alongside a signpost 'SINEU'.

• Co-occurrence based model (*Co-occur*): *Co-occur* is a very simple yet effective method for detecting relationship between words, we also take it as a baseline on the tag similarity calculation subtask instead of *LDA* method in subsection 2.3, and evaluate it in in the following part. Compared to our model, *Co-occur* is a weaker tool on detecting latent semantic information of tags.

Table 3. Result comparison with *F*1-value

	car	*girl*	*flower*	*bird*	*dog*	*sport*
W2V	0.217	0.531	0.233	0.389	0.325	0.442
Co-occur	0.683	0.724	0.631	0.557	0.626	0.702
Our model	0.881	0.767	0.854	0.871	0.873	0.801

We choose 6 kinds of frequently photographed images, which are 'car', 'girl', 'flower', 'bird', 'dog', 'sport', and respectively choose 100 images as the test datasets. For each image, we manually check the validity of every created image-knowledge link, which can help getting the precision easily for each model, and we take the union of all valid results of three different models as basis for getting recall of each model. Then the *F*1-value can be calculated and the results are shown in Table 3.

Two examples of experimental results are given in Fig.2 and Fig.3. Fig.2 shows an example of a semantic image ontology about buildings alongside a signpost 'SINEU', and not all links have been shown due to space limitation. Black solid lines show 'tag-knowledge' links we get directly from image tags, and red dotted lines indicate those links we get through knowledge extension function in our model. With our model, relation between Mallorca and Sineu, and relation between Palma and Sineu, can all be detected, and two knowledges 'dbr:Mallorca' and 'dbr:Palma' can be additionally linked to this image. For *predicates* as 'schema:indication' and 'schema:location', the 'schema' is short for <http://schema.org/>.

Fig.3 shows an example of a semantic image ontology about a Citroen car. With our model, knowledge 'dbr:Citroen_XM' can be easily detected with combination of tags 'citroen' and 'xm'. Besides, knowledge 'dbr:French' can be detected as a linked knowledge with *predicate* as 'schema:inLanguage', which is an error detection result. Actually, 'dbr:French' is for describing the country of the car brand 'dbr:Citroen', and the cause of this error is the shortage of *predicate* discriminate for different images.

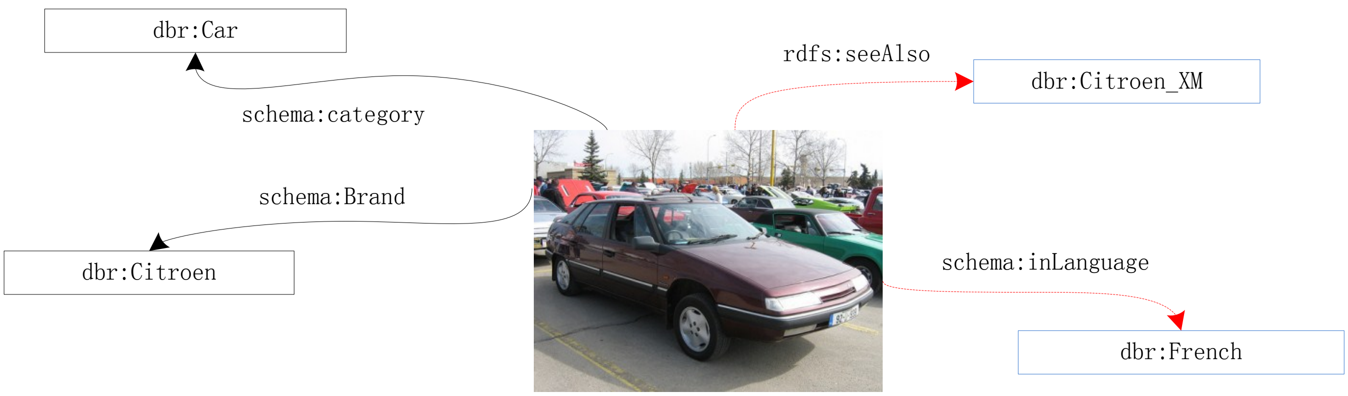

Figure 3. Example of a semantic image ontology (partial) about a Citroen car.

4. CONCLUSION AND FUTURE WORKS

In this paper, we address the problem of automatically linking images to semantic *KB* by using images' tagging information. Our framework lifted the *F*1-value of images' semantic link creation by 62.43%. However, this paper is only a preliminary work, and named entity disambiguation and named entity normalization will be considered in our next step work. Besides, *predicate* discriminate should be performed by considering adjacent tags as context. In addition, we will try to create ground-truth datasets and try some classification models on this task.

5. REFERENCES

[1] Arndt, R., Troncy, R., Staab, S., Hardman, L., Vacura, M. COMM: Designing a Well-Founded Multimedia Ontology for the Web. In: Proceedings of the 6th International Semantic Web Conference, 2007, pp. 30-43.

[2] Huang, R., Lu, S., Fotouhi, F. ImageSpace: an Image Ontology Creation and Annotation Tool for the Semantic Web. In: Proceedings of the 19th International Conference on Computers and Their Applications, 2004, pp.340-343.

[3] Ding, G., Xu, N. Automatic semantic annotation of images based on Web data. In: Proceedings of the 6th International Conference on Information Assurance and Security, 2010. pp. 317-322.

[4] Huiskes, M. J., Lew, M. S. The MIR Flickr Retrieval Evaluation. In: Proceedings of the 1st ACM SIGMM International Conference on Multimedia Information Retrieval, 2008, pp.39-43.

[5] Yang, S., Lu, W., Yang, D., Yao, L., Wei, B. Short Text Understanding by Leveraging Knowledge into Topic Model. In: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2015, pp.1232-1237.

[6] Phan, X-H., Nguyen, L-M., Horiguchi, S. Learning to classify short and sparse text & web with hidden topics from large-scale data collections. In: Proceedings of the 17th International World Wide Web Conference, 2008, pp.91-100.

[7] Naseer, A., Kume, T., Izu, T., Igata, N. LOD for All: Unlo-cking Infinite Opportunities, Semantic Web Challenge 2014.

[8] Cattuto, C., Benz, D., Hotho, A., Stumme, G. Semantic Analysis of Tag Similarity Measures in Collaborative Tagging Systems, LWA 2008: 18-26.