EXPLORING AND VISUALIZING TAG RELATIONSHIPS IN PHOTO SHARING WEBSITES BASED ON DISTRIBUTIONAL REPRESENTATIONS

Marie Katsurai and Miki Haseyama

Graduate School of Information Science and Technology, Hokkaido University
N-14, W-9, Kita-ku, Sapporo, Hokkaido, 060-0814, Japan
E-mail: katsurai@lmd.ist.hokudai.ac.jp, miki@ist.hokudai.ac.jp

ABSTRACT

This paper presents a method for exploring and visualizing tag relationships in photo sharing websites based on distributional representations of tags. First, we find a representative distribution of a tag, which is summarized by the mean and covariance, using features of tagged photos. This distributional representation can jointly consider the semantic meaning of tags and their abstraction levels. Then, based on the representative distributions, we derive two kinds of semantic measures on tag relationships. The extracted information is visualized in a graphical network to facilitate the understanding of tag usage. Experiments conducted using tagged photos collected from Flickr show that our tag network is more coherent to human cognition than other networks constructed by conventional methods.

Index Terms— tag relationship, knowledge extraction, visualization, photo sharing websites

1. INTRODUCTION

Recent years have witnessed the remarkable growth in the popularity of photo sharing websites: Flickr hosted over six billion photos on August 2011, and uploads have been increasing 20% per year over the last five years [1]. To efficiently manage a huge photo collection, these websites provide functions that allow users to annotate their uploaded photos with a set of keywords (called tags). Since tagging is performed by quite a number of users in a flexible way, there might be many tags that have the same or closely related meanings [2]. In order to facilitate the understanding of tag usage in the photo collection, exploring and visualizing semantic relationships between tags have recently attracted much research attention [2–12]. Most of the conventional methods represent each tag in a feature space as a single vector to measure inter-tag relatedness [8–12]. However, the vector-based representation cannot include information as to whether each tag has abstract or specific meanings. If the semantics of a tag are represented by a distribution with mean and covariance information, semantic relatedness between tags can be measured more accurately [13].

In this paper, we show how to explore semantic relationships between tags based on distributional representations of tags. In the proposed method, we find a representative distribution of each tag in a latent space estimated by using features from tagged photos. Since the distribution is summarized by the mean and covariance, we can implicitly represent not only semantic meaning of tags but also their abstraction levels. Based on the representative distributions, we derive two kinds of measures: the semantic relatedness between tags and the abstraction level of each tag. The two kinds of measures are effectively visualized in a novel graphical network. Experiments conducted on tagged photos collected from Flickr show that the tag network discovered by the proposed method is more coherent to human cognition than other networks constructed by conventional methods. The main contributions of this paper are two-fold: (i) our method can consider the abstraction levels for measuring semantic relatedness between tags by determining a representative distribution for each tag; and (ii) we visualize tag relationships by a novel graphical network that simultaneously shows the two kinds of measures: the abstraction levels of tags (as nodes) and the semantic relatedness between tags (as edges).

2. RELATION TO PRIOR WORK

Many methods so far have studied tag relationships in photo collections based on tag co-occurrence probabilities [2, 4–7]. On the other hand, to capture the semantic context of tags, some methods exploit tag co-occurrence vectors that reflect how many tags appear in the neighborhood of both tags [8–11]. This approach tends to obtain better results than the approach using only tag co-occurrence probabilities. In addition to these textual features, our previous work [12] considers that visual features are also effective cues to analyze semantic relationships, especially in a photo collection. Experimental results [12] showed that the collaborative use of tag co-occurrence vectors and visual feature vectors can improve the performance of measures on tag relationships over conventional methods that exploit single modality only. A common approach of our previous work and the conventional methods is that each tag is first represented as a single feature vector, and then the similarity between the vectors is computed as semantic relatedness between the tags. However, the vector representation cannot distinguish tags of different abstraction levels (for example, abstract or specific). The difference of abstraction level has shown effectiveness for measuring semantic relatedness in natural language processing research [14]. Thus, if a tag is modeled as one distribution with mean and covariance information, performance improvement of exploring tag relationships can be expected. This is the motivation of this paper.

The difference between our method and the conventional methods is not only how to explore tag relationships but also how to visualize them. The conventional methods mainly focus on displaying the calculated semantic relatedness between tags [2, 7, 8, 10, 11]. However, we consider that the abstraction level of a tag within the photo collection is important for effective tag-based photo search. Thus, in this paper, we simultaneously show the abstraction level of each tag and the semantic relatedness between tags as node weights and edge weights, respectively.

3. EXPLORING AND VISUALIZING TAG RELATIONSHIPS

This section presents a method for exploring and visualizing tag relationships in photo sharing websites based on distributional repre-

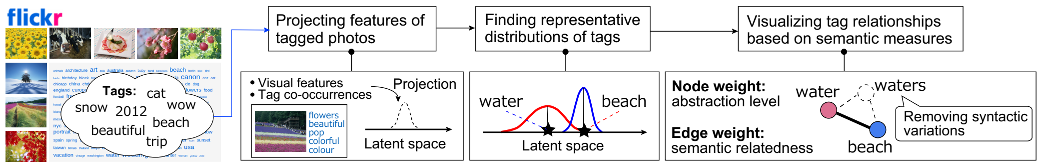

Fig. 1. An overview of the proposed method.

sentations of tags. An overview of the proposed method is shown in Fig. 1. As shown, we first extract tag co-occurrence vectors and visual feature vectors from all tagged photos and project them to a latent space (See 3.1). In the latent space, for each tag, we find a representative distribution with the mean and covariance information (See 3.2). By comparing the difference between the calculated distributions, we quantify two kinds of measures: semantic relatedness between tags and abstraction levels of tags. Finally, the extracted information is visualized in a novel graphical network (See 3.3).

3.1. Projecting features of tagged photos

Our previous work [12] showed that the collaborative use of tag co-occurrences and visual features can represent tag semantics better than the use of single modality in a photo collection. This paper also uses the two kinds of modalities to explore tag relationships. Given a tagged photo I_i ($i = 1, 2, \cdots, N$, where N is the number of photos), we first extract its d_x-dimensional visual feature vector $\boldsymbol{x}_i = [x_{i,1}, x_{i,2}, \cdots, x_{i,d_x}]^{\mathrm{T}}$ and d_y-dimensional tag co-occurrence vector $\boldsymbol{y}_i = [y_{i,1}, y_{i,2}, \cdots, y_{i,d_y}]^{\mathrm{T}}$. A pair $\{\boldsymbol{x}_i, \boldsymbol{y}_i\}$ is regarded as a sample in the proposed method. To discover the common underlying features on the different modalities, we use canonical correlation analysis (CCA) [15]. Specifically, under the probabilistic interpretation of CCA [16], we assume that features $\boldsymbol{x}$ and $\boldsymbol{y}$ are generated from the same latent variables z. According to [16], the posterior probability of the latent variables z given $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$ follows a normal distribution whose mean and variance are calculated by using the following equations:

$$\begin{aligned} \boldsymbol{m}_i &= E(z|\boldsymbol{x}_i, \boldsymbol{y}_i) \\ &= \begin{pmatrix} \mathbf{M}_x \\ \mathbf{M}_y \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} (\mathbf{I}-\boldsymbol{\Lambda}^2)^{-1} & -(\mathbf{I}-\boldsymbol{\Lambda}^2)^{-1}\boldsymbol{\Lambda} \\ -(\mathbf{I}-\boldsymbol{\Lambda}^2)^{-1}\boldsymbol{\Lambda} & (\mathbf{I}-\boldsymbol{\Lambda}^2)^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{U}_x^{\mathrm{T}}(\boldsymbol{x}_i - \bar{\boldsymbol{x}}) \\ \mathbf{U}_y^{\mathrm{T}}(\boldsymbol{y}_i - \bar{\boldsymbol{y}}) \end{pmatrix}, \quad (1) \\ \mathbf{S}_i &= var(z|\boldsymbol{x}_i, \boldsymbol{y}_i) \\ &= \mathbf{I} - \begin{pmatrix} \mathbf{M}_x \\ \mathbf{M}_y \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} (\mathbf{I}-\boldsymbol{\Lambda}^2)^{-1} & -(\mathbf{I}-\boldsymbol{\Lambda}^2)^{-1}\boldsymbol{\Lambda} \\ -(\mathbf{I}-\boldsymbol{\Lambda}^2)^{-1}\boldsymbol{\Lambda} & (\mathbf{I}-\boldsymbol{\Lambda}^2)^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{M}_x \\ \mathbf{M}_y \end{pmatrix}. \quad (2) \end{aligned}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix including the first q ($1 \leq q \leq \min\{d_x, d_y\}$) canonical correlations, $\mathbf{U}_x \in R^{d_x \times q}$ and $\mathbf{U}_y \in R^{d_y \times q}$ are projection matrices obtained by CCA, and $\mathbf{M}_x$ and $\mathbf{M}_y$ are arbitrary ($q \times q$) matrices with spectral norms smaller than one, such that $\mathbf{M}_x\mathbf{M}_y = \boldsymbol{\Lambda}$.

3.2. Finding representative distributions of tags

Let t_j ($j = 1, 2, \cdots, K$, K is the number of tags) be a tag in the target photo collections and Ω_j be a set of indexes of samples corresponding to t_j. In this subsection, we show how to find a representative distribution for each tag t_j using the sample set Ω_j. The i-th sample $\{\boldsymbol{x}_i, \boldsymbol{y}_i\}$ can be represented as a normal distribution whose parameters are calculated by Eqs. (1) and (2) in the latent space. Using Kullback-Leibler (KL) divergence as the distance measure between projected samples, we find the representative distribution with the mean $\boldsymbol{\mu}_j^*$ and covariance $\boldsymbol{\Sigma}_j^*$ by solving the following minimization problem:

$$p(z|\boldsymbol{\mu}_j^*, \boldsymbol{\Sigma}_j^*) = \arg \min_{p(z|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} \sum_{i \in \Omega_j} D_{KL}(p(z|\boldsymbol{m}_i, \mathbf{S}_i) || p(z|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)). \quad (3)$$

Note that the KL divergence between two normal distributions $p(z|\boldsymbol{m}, \mathbf{S})$ and $p(z|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ can be written as follows:

$$\begin{aligned} & D_{KL}(p(z|\boldsymbol{m}, \mathbf{S})||p(z|\boldsymbol{\mu}, \boldsymbol{\Sigma})) \\ = \ & \frac{1}{2}\Big\{ -q + tr(\mathbf{S}\boldsymbol{\Sigma}^{-1}) - \log |\mathbf{S}\boldsymbol{\Sigma}^{-1}| + (\boldsymbol{m}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{m}-\boldsymbol{\mu}) \Big\} \\ = \ & \frac{1}{2}\Big\{ B(\mathbf{S}, \boldsymbol{\Sigma}) + M_{\boldsymbol{\Sigma}^{-1}}(\boldsymbol{m}, \boldsymbol{\mu}) \Big\}, \quad (4) \end{aligned}$$

where $B(\mathbf{S}, \boldsymbol{\Sigma}) = tr(\mathbf{S}\boldsymbol{\Sigma}^{-1}) - \log |\mathbf{S}\boldsymbol{\Sigma}^{-1}| - q$ is the Burg matrix divergence and $M_{\boldsymbol{\Sigma}^{-1}}(\boldsymbol{m}, \boldsymbol{\mu}) = (\boldsymbol{m}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{m}-\boldsymbol{\mu})$ is the Mahalanobis distance parameterized by the covariance matrix $\boldsymbol{\Sigma}$. By using Eq. (4), the problem of finding the optimal representative distribution in set $\{p(z|\boldsymbol{m}_i, \mathbf{S}_i), i \in \Omega_j\}$ can be rewritten as follows [17]:

$$p(z|\boldsymbol{\mu}_j^*, \boldsymbol{\Sigma}_j^*) = \arg \min_{p(z|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} \frac{1}{|\Omega_j|} \sum_{i \in \Omega_j} \frac{1}{2}\Big\{ B(\mathbf{S}_i, \boldsymbol{\Sigma}_j) + M_{\boldsymbol{\Sigma}_j^{-1}}(\boldsymbol{m}_i, \boldsymbol{\mu}_j) \Big\}. \quad (5)$$

The second term can be viewed as minimizing the Bregman information [18] with respect to some fixed Mahalanobis distance. According to principle in Bregman clustering, this has a unique minimizer as follows [17]:

$$\begin{aligned} \boldsymbol{\mu}_j^* &= \frac{1}{|\Omega_j|} \sum_{i \in \Omega_j} \boldsymbol{m}_i, \quad (6) \\ \boldsymbol{\Sigma}_j^* &= \frac{1}{|\Omega_j|} \sum_{i \in \Omega_j} \Big\{ \mathbf{S}_i + (\boldsymbol{m}_i - \boldsymbol{\mu}_j^*)(\boldsymbol{m}_i - \boldsymbol{\mu}_j^*)^{\mathrm{T}} \Big\}. \quad (7) \end{aligned}$$

Thus, the representative distribution for tag t_j can be represented by a normal distribution parameterized by $\boldsymbol{\mu}_j^*$ and $\boldsymbol{\Sigma}_j^*$. This modeling of a tag can include both the mean and covariance information of the samples, which differs from the conventional vector-based tag representation [8–12].

3.3. Visualizing tag relationships based on semantic measures

Based on the distributional representations of tags, we derive two kinds of semantic measures: (1) semantic relatedness between tags and (2) abstraction levels of tags. We describe the details of the measures below.

1) Semantic relatedness between tags:

For measuring semantic relatedness from tag t_j to tag t_l ($l \neq j$), we calculate the KL divergence between their representative distributions as follows:

$$dist(t_j||t_l) = D_{KL}(p(z|\boldsymbol{\mu}_j^*, \boldsymbol{\Sigma}_j^*)||p(z|\boldsymbol{\mu}_l^*, \boldsymbol{\Sigma}_l^*)). \quad (8)$$

By using the KL divergence, we can consider both of the mean and scatter information for distribution differences [13]. Specifically, it can be expected that tags that not always co-occur but belong to the same category are effectively considered as related tags. Although the KL divergence is not a symmetric distance measure, its effectiveness on measuring similarities of distributions has been validated in several research fields [19, 20]. When considering both sides of the divergence, we can use Jensen-Shannon (JS) divergence as follows [19, 21]:

$$dist(t_j, t_l) = \frac{1}{2}\{dist(t_j||t_l) + dist(t_l||t_j)\}. \quad (9)$$

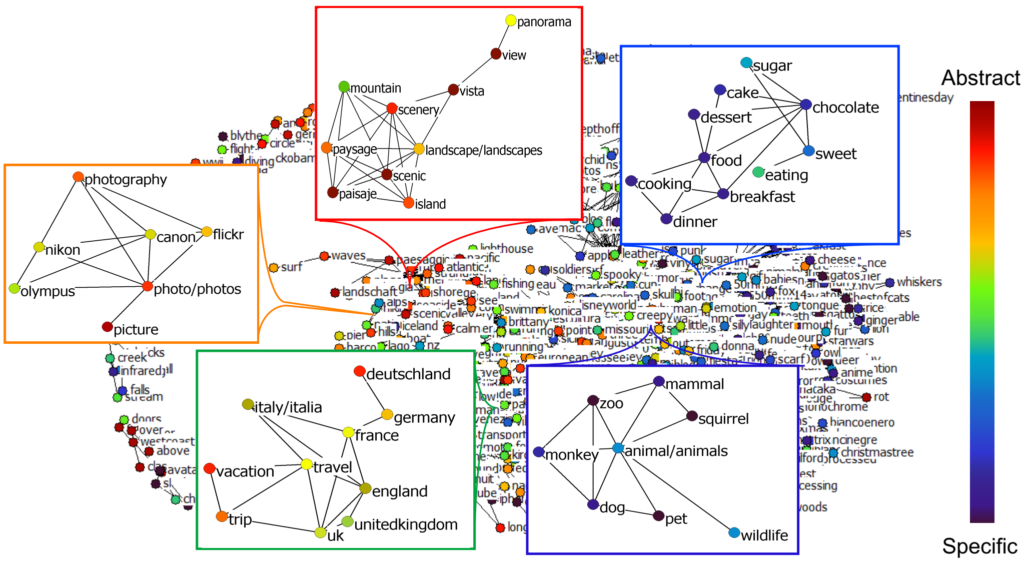

Fig. 2. Tag network constructed by the proposed method. Each node represents a tag and each edge between tags represents the strong semantic relatedness (only the top 1% edges are shown for easier viewing). Node color indicates the abstraction levels, which are set according to the color bar in the right side.

If $dist(t_j, t_l)$ is small, then it means that tags t_j and t_j are semantically related.

2) Abstraction levels of tags:

We assume that if a tag's samples are widely distributed in the latent space, then the tag has general and abstract meanings in the photo collection. Based on this assumption, we define the measure on abstraction levels of tags as follows:

$$\Phi(t_j) = \frac{1}{2}\left\{q + \log(2\pi)^q |\Sigma_j^*|\right\}. \quad (10)$$

This equation corresponds to the entropy of the representative distribution for tag t_j. A small $\Phi(t_j)$ means that samples corresponding to tag t_j are closely distributed to the centroid of those samples. On the other hand, a large $\Phi(t_j)$ means tag t_j is more abstract due to having several meanings. By monitoring the obtained scores, we can compare two tags in terms of abstraction levels and distinguish highly abstract tags and highly specific tags.

Finally, the calculated two kinds of semantic measures are represented as a graph $G = (V, E)$, where nodes $v_j \in V$ $(j = 1, 2, \cdots, K)$ correspond to tag t_j, and the edges $e(v_j, v_l) \in E$ $(j, l = 1, 2, \cdots, K)$ indicate the semantic relationships between tags t_j and t_l. In graph G, we provide the semantic relatedness $dist(t_j, t_l)$ as the weight of edge $e(v_j, v_l)$, and the abstraction level $\Phi(t_j)$ as the weight of node v_j. Due to the nature of tagging [22], there may be a number of syntactic variations such as misspellings (e.g., "catarog", catalog) and the use of plural or singular forms (e.g., building, buildings). In order to remove syntactic variations in the visualization, we employ the normalized Levenshtein distance [10] as a string similarity metric. If the edge $e(v_j, v_l)$ has high semantic relatedness with low normalized Levenshtein distance, then the edge is removed from the network and the corresponding nodes v_j and v_l are integrated for easier viewing. The constructed network can facilitate the understanding of tag usage in a photo collection.

4. EXPERIMENTS AND VISUALIZATIONS

In this section, we show experimental results to verify the effectiveness of the proposed method. Our experiments are conducted on a MIRFLICKR-1M dataset [23], which contains 1,000,000 tagged photos collected from Flickr. All tags that are used by less than 50 users or whose samples are less than 700 have been filtered, and the remaining 1,915 tags are used for performance evaluation. In the experiments, we use 3,575-dimensional tag co-occurrence vectors and 2,986-dimensional visual feature vectors based on RGB-SIFT [24]. We extract the tag relationships by the proposed method from the dataset and depict the network by using NetDraw [25]. Since it is difficult to show the whole network, we show only the top 1% of strongest relatedness as edges for easier viewing. The depicted network is shown in Fig. 2, where node colors are set according to node weights by using the color bar shown in the right side. Red nodes indicate the corresponding tags are at high levels of abstraction (i.e., abstract tags), whereas blue nodes indicate the corresponding tags are at lower levels of abstraction (i.e., specific tags). In the enlarged parts of Fig. 2, we can find that tags such as "photo" and "picture" are

Table 1. Results of measuring abstraction levels of tags, in which the top 12 abstract and the top 12 specific tags are shown. Some specific tags are explained in italics.

Top Abstract Tags	Rank	Top Specific Tags
goldenphotographer	1	origami (*paper art*)
goldenglobe	2	bento (*lunch box in Japanese*)
experiment	3	dragoncon (*cosplay event*)
pictureperfect	4	blythe (*fashion doll*)
thebestofday	5	neko (*cats in Japanese*)
gt	6	gatos (*cats in Spanish*)
digitalcameraclub	7	bestofcats
multimegashot	8	secondlife (*virtual world*)
isawyoufirst	9	squirrel
sensational	10	tabby
international	11	crochet
proudshopper	12	katze (*cats in German*)

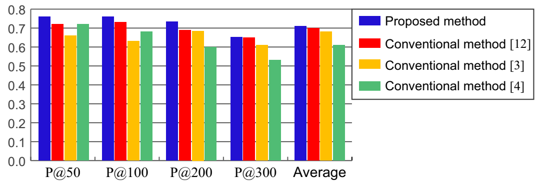

Fig. 3. Precisions of abstraction levels of tags.

Table 2. User assessments of the semantic relatedness measured by the proposed method and the conventional method.

Proposed method		Conventional method [11]	
Tag pair	Score	Tag pair	Score
newyork - nyc	4.78	aircraft - plane	4.64
35mm - film	4.77	election - president	4.58
blackwhite - bw	4.69	aviation - aircraft	4.56
ocean - sea	4.61	fantastic - stunning	4.54
boy - child	4.54	fantastic - excellent	4.53
plants - garden	4.52	stunning - gorgeous	4.49
≀	≀	≀	≀
365days - me	3.89	do - janeiro	3.78
photomatix - hdr	3.84	delete - 10	3.73

considered as highly abstract tags. On the other hand, tags such as "squirrel" and "chocolate" are considered to have relatively specific meanings. In the following subsections, we evaluate our network in terms of abstraction levels used as node weights (see 4.1) and semantic relatedness used as edge weights (see 4.2), respectively.

4.1. Evaluation of abstraction levels

First, we investigate what kinds of tags are assumed to be abstract or specific tags by the proposed method. Table 1 lists the top 12 abstract tags and the top 12 specific tags based on abstraction levels measured by the proposed method. In this table, we can find that the top specific tags actually represent specific objects, scenes, or activities, while the top abstract tags represent opinions or impressions.

To quantitatively evaluate the performance, we manually divide all tags into objects, scenes, activities, and others in the same way as conventional works [3, 12]. As an evaluation metric, we calculate the precision obtained as the top m specific tags as follows:

$$P@m = \frac{\#\text{tags divided into either of objects, scenes, or activities}}{\text{the top specific } m \text{ tags estimated by the method}}. \quad (11)$$

If $P@m$ approaches 1, it means the method can accurately locate the specific tags in the lower levels. We also apply conventional methods that estimate abstraction levels of tags in the photo collection [3, 4, 12] for performance comparison. The results are shown in Fig 3. As shown in this figure, the proposed method can place the tags representing objects, scenes, or activities in a lower abstraction level than other methods. These abstraction levels are implicitly considered for measuring semantic relatedness between tags as covariance information.

4.2. Evaluation of semantic relatedness

We now evaluate semantic relatedness between tags which are computed as edge weights of the network. It is a non-trivial task to quantitatively evaluate the results of semantic relatedness between tags due to the lack of ground truth. One obvious metric for evaluating extracted knowledge is its correspondence with human judgment [26]. For subjective evaluation, we present 32 users a list of

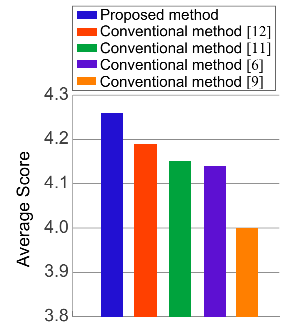

Fig. 4. Average user scores.

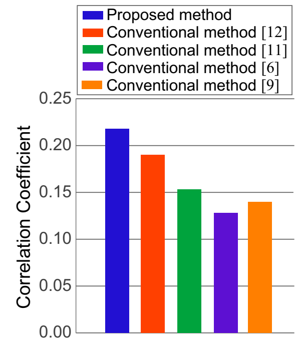

Fig. 5. Calculated correlation to WordNet.

the top 100 related tag pairs calculated by each method. For each tag pair, the user is required to give a score ranging from 1 ("unrelated") to 5 ("highly related") according to their knowledge. Note that we inject 10 pairs with appropriate scores into the list in order to discard poor-quality judgment. The mean of the scores from all users is considered as the "user score" for the tag pair. Table 2 shows some of the pairs extracted by the proposed method and the conventional method [11]. Furthermore, we average user scores provided for the top 100 tag pairs extracted by each method. The average scores of different methods are shown in Fig. 4, where our method obtains the best user scores.

For more extensive experiments, we use WordNet [27] as ground truth. Although WordNet only contains a small portion of the tags that are used in Flickr, it can be a reasonable benchmark based on human cognition. The number of tag pairs used in this experiment is 41,690. We sort the tag pairs according to the semantic relatedness for each method and compare their rankings with WordNet rankings by using the Spearman's correlation coefficients. If the two rankings are exactly same, the coefficient between these two rankings is 1. The calculated Spearman's correlation coefficients are shown in Fig. 5. From this comparison, we can find that our method outperforms conventional methods in measuring semantic relatedness. This improvement is based on the distributional representation of tags, which can jointly consider the mean and covariance information. In future work, we will conduct more experiments based on a larger dataset.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we present a method for exploring tag relationships in photo sharing websites based on distributional representations of tags. The extracted relationships are visualized in a novel graphical network. Experiments conducted on one million tagged photos collected from Flickr show that tag network discovered by the proposed method is more coherent to human cognition than other networks constructed by conventional methods. In future work, we will investigate tag ambiguity by assigning more than two representative distributions to a single tag. Furthermore, we will apply the tag network to tag recommendation and photo search.

6. ACKNOWLEDGEMENTS

This work was partly supported by Grant-in-Aid for Scientific Research on Innovative Areas 24120002 from the MEXT.

7. REFERENCES

[1] K. Kremerskothen, "Flickr blog: 6,000,000,000," `http://blog.flickr.net/en/2011/08/04/6000000000/`, Aug 2011, Accessed: 25/11/2012.

[2] Y. Hassan-Montero and V. Herrero-Solana, "Improving tag clouds as visual information retrieval interfaces," in *Proc. Int. Conf. Multidisciplinary Information Sciences and Technologies (InSciT)*, 2006.

[3] A. Sun and Sourav S. Bhowmick, "Quantifying tag representativeness of visual content of social images," in *Proc. Int. Conf. Multimedia (MM)*, 2010, pp. 471–480.

[4] P. Heymann, D. Ramage, and H. Garcia-Molina, "Social tag prediction," in *Proc. Int. Conf. Research and development in information retrieval (SIGIR)*, 2008, pp. 531–538.

[5] R.L. Cilibrasi and P.M.B. Vitanyi, "The google similarity distance," *IEEE Trans. Knowledge and Data Engineering*, vol. 19, no. 3, pp. 370 –383, Mar. 2007.

[6] D. Liu, X. S. Hua, L. Yang, M. Wang, and H. J. Zhang, "Tag ranking," in *Proc. Int. Conf. World wide web (WWW)*, 2009, pp. 351–360.

[7] G. Begelman, P. Keller, and F. Smadja, "Improving search and exploration in the tag space," in *Proc. Collaborative Web Tagging Workshop at WWW*, 2006.

[8] L. Specia and E. Motta, "Integrating folksonomies with the semantic web," in *Proc. European conference on The Semantic Web: Research and Applications (ESWC)*, 2007, pp. 624–639.

[9] C. A. Yeung, N. Gibbins, and N. Shadbolt, "Contextualising tags in collaborative tagging systems," in *Proc. Int. Conf. Hypertext and hypermedia (HT)*, 2009, pp. 251–260.

[10] J. W. van Dam, D. Vandic, F. Hogenboom, and F. Frasincar, "Searching and browsing tag spaces using the semantic tag clustering search framework," in *Proc. Int. Conf. Semantic Computing (ICSC)*, 2010, pp. 436 –439.

[11] A. Ksibi, A. Ben Ammar, and C. Ben Amar, "Effective concept detection using second order co-occurence flickr context similarity measure SOCFCS," in *Proc. Int. Workshop on Content-Based Multimedia Indexing (CBMI)*, 2012, pp. 1 –6.

[12] M. Katsurai, T. Ogawa, and M. Haseyama, "A cross-modal approach for extracting semantic relationships of concepts from an image database," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, 2012, pp. 2373–2376.

[13] B. Chen, W. Lam, I. Tsang, and T. L. Wong, "Location and scatter matching for dataset shift in text mining," in *Proc. Int. Conf. Data Mining (ICDM)*, 2010, pp. 773 –778.

[14] C. Leacock and M. Chodorow, "Combining local context and wordnet similarity for word sense identification," in *WordNet: An electronic lexical database*, C. Fellbaum, Ed., pp. 265–283. MIT Press, 1998.

[15] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, 1936.

[16] F. R. Bach and M. I. Jordan, "A probabilistic interpretation of canonical correlation analysis," Tech. Rep. 688, Department of Statistics, University of California, Berkeley, 2005.

[17] J. V. Davis and I. Dhillon, "Differential entropic clustering of multi-variate Gaussians," in *Neural Information Processing Systems (NIPS)*. 2007, pp. 337–344, B. Scholkopf, J. Platt, and T. Hoffman, Eds. Cambridge, MA: MIT Press.

[18] A. Banerjee, S. Merugu, I. S. Dhillon, and J. Ghosh, "Clustering with bregman divergences," *Journal of Machine Learning Research*, vol. 6, pp. 1705–1749, Dec. 2005.

[19] W. Wang, P.M. Barnaghi, and A. Bargiela, "Probabilistic topic models for learning terminological ontologies," *IEEE Trans. Knowledge and Data Engineering*, vol. 22, no. 7, pp. 1028–1040, July 2010.

[20] R. Miotto and N. Orio, "A probabilistic model to combine tags and acoustic similarity for music retrieval," *ACM Trans. Information Systems*, vol. 30, no. 2, pp. 8:1–8:29, May 2012.

[21] N. Rasiwasia, P.J. Moreno, and N. Vasconcelos, "Bridging the gap: Query by semantic example," *IEEE Trans. Multimedia*, vol. 9, no. 5, pp. 923 –938, Aug. 2007.

[22] M. E. I. Kipp and D. G. Campbell, "Patterns and inconsistencies in collaborative tagging systems: An examination of tagging practices," *Proc. American Society for Information Science and Technology (ASIST)*, vol. 43, no. 1, pp. 1–18, 2006.

[23] M. J. Huiskes and M. S. Lew, "The MIR flickr retrieval evaluation," in *Proc. ACM Int. Conf. Multimedia information retrieval (MIR)*, 2008, pp. 39–43.

[24] K.E.A. van de Sande, T. Gevers, and C.G.M. Snoek, "Evaluating color descriptors for object and scene recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1582–1596, Sept. 2010.

[25] S. P. Borgatti, "Netdraw software for network visualization," Analytic Technologies: Lexington, KY, 2002.

[26] S. Patwardhan, "Using wordnet-based context vectors to estimate the semantic relatedness of concepts," in *Proc. EACL 2006 Workshop on Making Sense of Sense: Bringing Computational Linguistics and Psycholinguistics Together*, 2006, pp. 1–8.

[27] George A. Miller, "WordNet: a lexical database for english," *Commun. ACM*, vol. 38, no. 11, pp. 39–41, Nov. 1995.