2012 IEEE International Conference on Systems, Man, and Cybernetics
October 14-17, 2012, COEX, Seoul, Korea

Improving the Quality of Tags Using State Transition on Progressive Image Search and Recommendation System

Wen-Hau Du, Jer-Wei Rau, Jen-Wei Huang*, Yung-Sheng Chen
Department of Electrical Engineering, Yuan Ze University, Taoyuan, Taiwan
* Department of Electrical Engineering, National Cheng Kung University, Tainan, Taiwan
s9746531@ee.yzu.edu.tw, joe@ee.yzu.edu.tw, jwhuang@mail.ncku.edu.tw, eeyschen@saturn.yzu.edu.tw

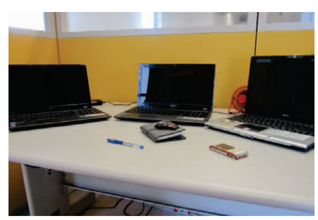
***Abstract*—In recent years, tagging methods have been widely used on the Web for multimedia search and recommendation systems. Social tags are the keywords annotated by users to the multimedia objects. These tags contain the information which allows multimedia objects to be located and classified. Unfortunately, irrelevant information and noise are frequently included in such tags, especially when we use social tagging systems. Imperfect as they may be, social tags are still a significant source of human-generated contextual knowledge. We present the Waking And Sleeping (WAS) algorithm to overcome the obstacle of noisy tags. WAS refines social tags of each image and assigns each tag into either the waking state or the sleeping state. With WAS, the recommendation system not only enhances the accuracy of the recommendation results based on tags with higher quality but also reduces the comsuption of computational resources. The experimental results demonstrate that the proposed approach is superior to compared methods.**

***Index Terms*—Tag; Image Annotations; Tag Quality Improvement; Recommendation System**

I. Introduction

The advent of digital cameras, cheap storage, and communication technology has paved the way for the exponential growth in the amount of multimedia content generated by Internet users. Several prominent examples of this type of growth include Flickr[1], YouTube[2], and Facebook[3]. According to statistical reports in 2011, the quantity of image uploading on Photobucket[4] was more than 9 billion. Google Picasa[5] reported 7 billion images uploaded, whereas Flickr had more than 5 billion images. To search or find the most relevant images using large amounts of data sets has become a considerable challenge for users, and thus merits discussion. With the advancements of Web 2.0 technology, users can easily create multimedia content and most of the social medium repositories allow people to annotate their content. In a social tagging system, tags can be combined with those created by other users to form a collective body of social tags. A collection of the adequately large set of tags is built up by many people over time. Such a collection of tags can be used to retrieve for known items and explore new items. Recently, the performance of Web image search and recommendation system depends particularly on the quality of the tags.

Image search and recommendation systems must consider the relationships between tags and images in databases to provide the most suitable results to users. However, numerous noisy tags seriously affect the accuracy of the results. According to Sigurbjörnsson [1], many users enjoy using different types of personal information to annotate their images. However, only approximately 50% of social tags are actually related to the images. For example, the image in Fig. 1 was queried on Flickr with the keyword "notebook". Approximately, 4 to 5 tags accurately describe the context of the given image, but the others are bad tags because these tags contain only the personal significance ("20110731", "21500 NT", "happy", "Johns"), the irrelevant information ("booty", "VS"), or typos ("Natebook", "Myroom").



Notebook	Johns'
Mobile	office
Natebook	NBs
20110731	21500 NT
Myroom	booty
VS	happy

Fig. 1. The example randomly downloaded from Flickr and its tags.

According to Wu [2], only 40% of tags in the Flickr database are with high quality. Those tags with lower quality not only reduce the accuracy of the search and recommendation results but also burden the search and recommendation system with heavy computation resources. Although some semantic-based image search and recommendation engines exhibit effective self-improving potential, there are some drawbacks remain. First, many image search engines still suffer from the un-controlled styles of noisy tag. Consequently, many images and tags may not be effectively related to one another. Second, the large number of noisy tags costs image search engines considerable computing time. Third, those words are frequently used as keywords within a short period of time are not considered as more important tags.

[1] Flickr, http://www.ickr.com/
[2] YouTube, http://www.youtube.com/
[3] Facebook, http://www.facebook.com/
[4] Photobucket, http://photobucket.com/
[5] Google Picasa, https://picasaweb.google.com/

978-1-4673-1714-6/12/$31.00 ©2012 IEEE

In the previous research, Liu [3] proposed a framework to improve badly annotated tags associated with images. Liu adopts a pre-filtering process for all tags according to the information in Wikipedia[6]. Hence, we temporarily name it WTQI, Wikipedia-based Tag Quality Improvement. WTQI filters out some bad tags and adds several tags based on the coherence of visual similarity and semantic similarity of images. We can expand the good quality of tags in an image using this method. However, WTQI not only removes valuable tags, but also introduces several irrelevant tags. Huang [4] proposed PISAR algorithm to progressively enhance the accuracy of search and recommendation results by collecting records of user interactions in the system. Based on the clicked frequency of tags and images are extremely helpful for calculating relationships between each other. However, PISAR algorithm also causes an inequitable side effect. Because of the clicked frequency of the older tags are more than those of the newer tags. Even if the new tags are more valuable than the older tags for an image, newer tags will not be recommended to users.

In this work, we propose a Waking And Sleeping (WAS) algorithm to remedy these problems. Fig. 2 shows that WAS assigns tags into either the waking state or the sleeping state to improve the quality of tags in an image. With the help of WAS, the recommendation system temporarily ignores those tags in the sleeping state, and therefore the recommendation results will be more accurate according to tags with higher quality. Consequently, the computational resources on the similarity calculation are reduced. There are two modules in WAS, Importance Filter and Relationship Retriever. Importance Filter not only counts the tag click events but also regards the time interval of the click events. Relationship Retriever recycles the useful tags from the sleeping state candidate, because of those tags are very similar to the tags in the waking state.

The importance filter measures how important a tag is to an image in a social tagging system. The importance weight increases proportionally to the number of a tag click event appears in an image, but is offset by the frequency of the tag in the social tagging system. Nevertheless, the importance weight of the new tag which frequency growths within a short period may still be smaller than the importance weight of old tag. Therefore, we consider the significance of the frequency and also the combination of the value of time. Applying this formula, we can get the probability of the tag used by users of an image during a particular period.

We observe that some of the significant but especial tags are assigned to sleeping state candidate. For Example, the formal name of a long-hair rabbit is the Angolan Rabbit. Angolan Rabbit will be assigned to sleeping state candidate, because it is an uncommon name for daily usage. Therefore, we apply the relationship retriever to reassign the valuable tags to waking state from sleeping state candidate. Relationship Retriever applies the cosine similarity to measure each tags in sleeping state candidate with the tags in waking state. If the value is higher than the threshold, the tag is valuable for the image and assign the tag to the waking state.

[6]Wikipedia, http://en.wikipedia.org/

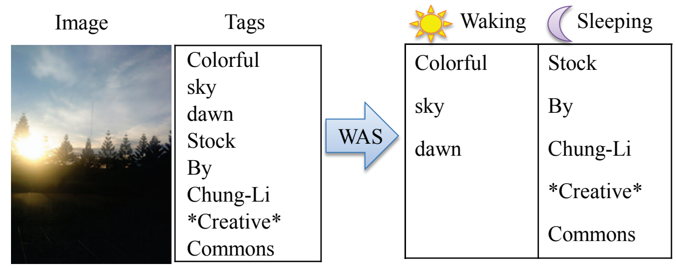


Fig. 2. The concept of Waking And Sleeping algorithm.

The remainder of this paper is organized as follows. Section 2 briefly introduces related studies and background on social tagging systems. Section 3 provides a detailed view of WAS algorithm. The experimental results and discussion are presented in Section 4. Finally, Section 5 concludes this paper.

II. Related works and background

A. Related works

With the Web 2.0 technology, tagging becomes a popular function to annotate multimedia items on the Web. A previous study [5] showed that the social incentives for tagging appear to be the essential motivations in motivating users to annotate multimedia items on the web. With this advantage, the semantic-based image search and recommendation systems are easily accomplished with that collective knowledge from the social tags. Even if social tags contain many bad tags, social tags are still the good sources of human-generated collective knowledge about the multimedia items on the web. Therefore, many research groups have begun to study on social tagging clustering models for large-scale web image retrieval. However, only few of previous studies have focused on the problem of the social tags with bad quality and can be generally divided into two categories: the content-based methods and the semantic-based methods.

The content-based methods aim to annotate images automatically or semi-automatically using low-level image features [6, 7]. A previous study [8] proposed a novel semi-auto image tagging technique for the images on Flickr. With the spatial pyramid matching classifier, the sparse coding algorithm measures the low-level features of each images to classify the relevant images. Social tags are re-ranked according to the relevant images, and the top-ranking tags are assigned for the new image. With semi-auto image tagging technique, the system annotates images automatically and reduces the noisy level of tags.

The studies [9, 10] on the semantic relationships between the concepts in the image have also been a popular research theme recently. Xu [11] argued that the criteria for good tags should have the following five properties: high coverage of multiple facets, high popularity, minimal effort, uniformity (normalization), and exclusion of certain types of tags. Those previous studies have focused on social tags cluster analysis, classification, and categorization. Several works [12, 13] have compared social tags with the concepts formally defined in WordNet[7]. WordNet is a large semantic lexicon database for English language. In WordNet, the nouns, the verbs, the adjectives, and the adverbs are organized into synonyms for each network, and each synonym set represents a basic semantic concept. Numerous scholars have used synonym sets as a foundation to determine the quality of a tag.

Liu [3] proposed WTQI to improve the tag sets in an image by simultaneously modeling the consistency of visual and semantic similarities with the compatibility of tags before and after improvement. At the first, Liu adopts a pre-filtering process for all tags in the database according to the information in Wikipedia, and filters out those tags with bad quality. Because the Wikipedia is a taxonomy system and provides expert annotations, the pre-filtering process ensures the quality of social tags. With these tags with good quality, WTQI adds several tags to an image based on the coherence of Content-based similarity and semantic-based similarity. To measuring content-based similarity requires substantial calculating time deal with the low-level feature of the images in the database. In addition, there are still several unrelated tags given to an image after WTQI.

B. PISAR system

PISAR is a search and recommendation system on the Web that can progressively enhance the accuracy of search and recommendation result based on users' interaction records. PISAR system presumes the interactions between the user and the search system is significant when the user attempts to retrieval a particular type of image. The click events after users query contain lot of information on their motivation that is extremely helpful for constructing relationships between tags and images.

In addition, we discovered that the usage counts of bad tags and good tags in PISAR system varies considerably. Many user retrieval same kind of images according to the similar tags. The click events are similar for the query results and the recommendation results. This indicates that many users select the image according to the tag when the clicked frequency is higher. It means that the tag contributes highly to the image. The clicked counts of tags are extremely helpful for calculating relationships between tags and images. PISAR discover the relationships between tags and images follow the concept of co-occurrence with cosine similarity method.

III. Waking and sleeping state

A. Definition and Properties

The waking state and the sleeping state are defined as follows:

1) Waking state: The tags in the "Waking state" have sufficient strength to represent the image because these tags are highly relevant to the image and are often used by users within a period of time. The tags in the waking state not only boost the accuracy of the recommendation results but also reduce the computational resources on the similarity calculating.

2) Sleeping state candidate: The "Sleeping state candidate" is the middle stage of WAS algorithm. The "Importance Filter" assigns the tags with good quality to the "Waking state" and the rest of tags are assigned to the "Sleeping state candidate". But there are still part of the tags in the "Sleeping state candidate" provide usable information for the image. We measure the similarity of tags between the "Sleeping state candidate" and the "Waking state" then reassign the tags.

3) Sleeping state: The tags in the "sleeping state" contain only personal significance or typos and not often be used. These tags do not participate in the similarity calculation with other tags during the image recommendation and tag recommendation processes.

B. WAS Algorithm

We propose WAS algorithm for assigning social tags to the waking state or the sleeping state to improve the quality of social tags. The flow chart of WAS algorithm is shown in Fig. 3. WAS algorithm uses two modules, the importance filter and the relationship retriever, to assign tags to either waking state or sleeping state.

First, we propose the importance filter to determine the importance value of a tag for an image and assign tags to either the waking state or the sleeping state candidate. The importance filter incorporates the concept of click counts within a period of time as an importance value to prevent the phenomena of the stronger tags are getting strong and the newer tags are getting weak.

After applied with the importance filter, tags are assigned to waking state and sleeping state candidate. We observe that some of the significant but especial tags are assigned to sleeping state candidate. For Example, the formal name of a long-hair rabbit is the Angolan Rabbit. Angolan Rabbit will be assigned to sleeping state candidate, because it is an uncommon name for daily usage. Therefore, we design the relationship retriever to pick valuable tags from the sleeping state candidate back to the waking state and the rest of tags are assigned to the sleeping state. The relationship retriever uses cosine similarity to calculate the tags relationship between the waking state and the sleeping state candidate. If the value of the relationship is higher than the threshold, the relationship retriever picks the tag in the sleeping state candidate up to the waking state. Otherwise, the tags in the sleeping state candidate are assigned to the sleeping state.

Two modules in WAS algorithm are described below:

1) Importance Filter: The importance filter estimates the importance value, IMP_{ij} of the tag t_i in an image I_j. The importance weight IW_{ij} similar to TF-IDF [14] and the ACT_{ij} involves considering the activity degree of the tags in an image during a particular period.

$$IMP_{ij} = IW_{ij} * ACT_{ij} \tag{1}$$

[7] WordNet, http://wordnet.princeton.edu/

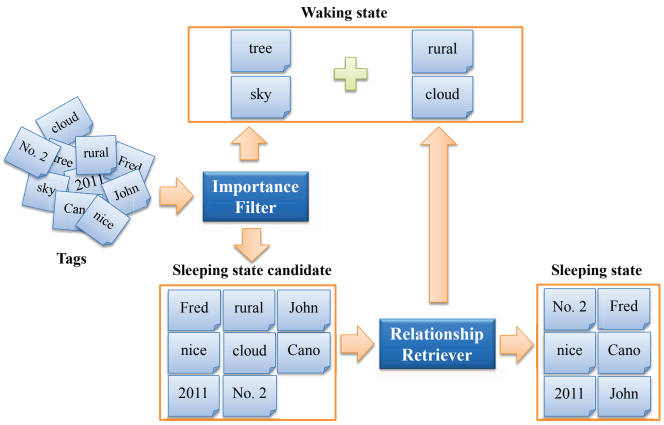


Fig. 3. Flow chart of WAS algorithm.

The Threshold $Threshold_j$ for each image are used to filter out unimportant tags.

$$Threshold_j = AVG(\sum_{t_i \in I_j} IMP_{ij}) \tag{2}$$

If IMP_{ij} is lower than the $Threshold_j$, the tag t_i will be assigned to the sleeping state candidate.

IW_{ij} is a statistical measurement used to evaluate importance of a tag in the social tagging database. Like TF-IDF, we regard tags as terms and images as documents. The importance weight of the tag t_i in an image I_j, can be denoted as:

$$IW_{ij} = \frac{n_{ij}}{\sum_i n_{ij}} log \frac{N}{N_i} \tag{3}$$

where $\sum_i n_{ij}$ denotes the total number of tags in an image I_i, and n_{ij} is the number of occurrences of the tags in an image. N_i is the number of images containing the tag t_i, and N is the number of images in the database. The importance weight IMP_{ij} is the critical factor for the tag t_i. Using importance weight, we can extract the critical factor of the tag from user interaction records.

How to emphasize such tags that have the high interaction frequencies within a short period of time is importance crucial key for this stage. For example, the Bumblebee Transformer from the Chevrolet Camaro sports car is the most well-known. User may want a recommendation on Bumblebee when they search for a Chevrolet Camaro. However, if we use "Chevrolet Camaro" as a keyword to search on Web before the movie. After the movie is released, people who use "Chevrolet Camaro" as a keyword will locate "Bumblebee" instantaneously within a short period.

Nevertheless, the importance weight of the new tag which frequency growths within a short period may still be smaller than the importance weight of old tag. Therefore, we consider the significance of the frequency and also the combination of the value of time. The second part of the weight calculation involves considering the activity degree of the tag in the image during a particular period.

$$ACT_{ij}(k, \lambda) = \frac{e^{-\lambda}\lambda^k}{k!} \tag{4}$$

Where e is the base of the natural logarithm, k is the number of occurrences of a click event, such as selecting an image or clicking a tag from a tag list. λ is a positive real number, equal to the expected number of occurrences that occur during a particular period. Applying the ACT formula, we can get the probability of the tag used by users of an image during a particular period.

2) Relationship Retriever: At this stage, we recycle useful tags from the sleeping state candidate that are very similar to the tags in the waking state. The cosine similarity is widely used in text mining. We adopt the cosine similarity to measure the correlations between entire waking tags and all sleeping state candidate tags in an image. In this work, the tag t_i vector can be denoted as

$$t_i = \{I_1, I_2, \cdots, I_n\} \tag{5}$$

where the image I_i represents the elements in the tag vector, if the tag exists in ith image the element value is 1; otherwise, it is 0. According to the cosine similarity, we can obtain the similarity between tag in the waking state t_p and tag in the sleeping state candidate tag_q as follows:

$$Similarity(\vec{t_p}, \vec{t_q}) = \frac{\vec{t_p}\vec{t_q}}{\|t_p\|\|t_q\|} \tag{6}$$

The minimum threshold for an image is given by

$$AVG(Similarity(\vec{t_p}, \vec{t_q})) \tag{7}$$

If the cosine similarity between t_p and t_q is higher than the minimum threshold, then it means a strong relationship exists between t_p and t_q. Therefore, "relationship retriever" reassign the tag from the sleeping state candidate to the waking state, and the rest of the sleeping state candidate tags will be assigned as the sleeping state.

Furthermore, WAS algorithm is designed only for the results of recommendation, not affecting the search results. The image search engines still can retrieval particular images using bad quality of tags. With the WAS algorithm, search and recommendation system can extract the higher quality of tags for each image. The higher quality of tags enhance the accuracy of the recommendation results and reduce many computational resources on calculating the similarity of the sleeping tags.

IV. Experiment

A. Experimental Environment

We implement WAS algorithm based on PISAR system. The data set of PISAR contains 7,000 images and nearly 20,000 tags. Those images were randomly downloaded from Flickr and divided into 14 categories. They are scenery, landscape, landmark, night, street, building, river, lake, beach, mountain, forest, sunset, sunrise and sky. Each category contains 500 images.

TABLE I
A PART OF TAG LIST

Rank	Tag	Rank	Tag	Rank	Tag
730	canal+	739	novi	748	cyan
731	mtv	740	rio	749	ave
732	star people	741	wing	750	oc
733	chiarappa	742	ridge	751	nile
734	peintre	743	iowa	752	wy
735	deco	744	200308c	753	beam
736	1	745	twinlakes	754	tete
737	a.d.	746	sd	755	kule
738	essex	747	400	756	s700

To prove that using WAS algorithm can improve the quality of tags in an image and enhance the recommendation results, we compare experimental results with [4]. Because PISAR system does not improve the quality of tags, we compare WAS algorithm with WTQI [3]. In the experiments, we calculated the similarity using the three methods, PISAR system, WAS-PISAR, and WTQI every day. We selected ten changeless tags and ten changeless images to observe the mutative situation every day. Finally, we randomly chose thirty users to give the experimental results comments.

B. Data Analysis on PISAR

Fig. 4 shows that most of the images in the PISAR database have the 3 to 11 tags, with the X-axis being the number of tags in each image and the Y-axis being the number of images. Furthermore, we sorted all the tags according to frequency and discovered many irrelevant tags about the top 25% of the sorting list, as shown in TABLE I. These bad qualities of tags have the high frequency, but they are meaningless to most of users. Therefore, we should prevent such bad quality of tags as the recommendation results.

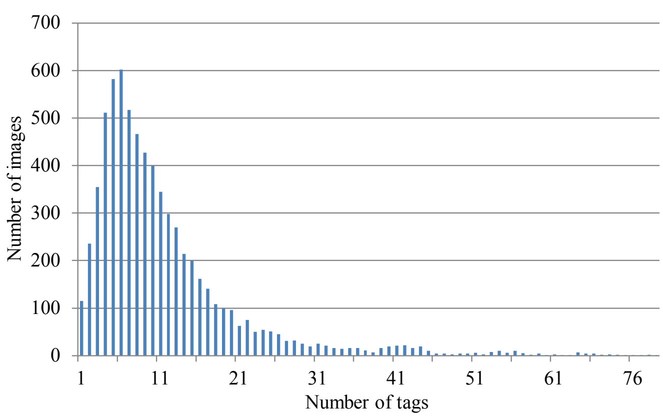


Fig. 4. Distribution of the tags in PISAR.

C. Quality of Tags

Fig. 5 shows an image with five tags that are annotated by users. The tags, "sky" and "cloud", are directly related to this image. The "beautiful" tag is classified into the type of concept. The "nyc" and "holiday" tags illustrate the location and time information.

WTQI filters out the "nyc" tag and also refines the tag vectors using the similar images. Several new tags such as "beach", "water", "sea", and "ocean" are added in this image based on this issue. But the impression descriptions and the location abbreviations are dropped.

PISAR and WAS advance the accuracy of recommendation results by analysis the user interaction data. WAS is based on PISAR, we name it "WAS-PISAR". WAS-PISAR shows the most excellent result for this case by simply allocate those tags into waking and sleeping results.

PISAR	WTQI	WAS-PISAR
sky, cloud, nyc, holiday, beautiful	beach, water, sky, sunset, sunrise, sea, clouds, ocean, holidays, river	Waking state: sky, cloud
		Sleeping state: nyc, holiday, beautiful

Fig. 5. An image and their tags after quality improvement.

D. Tag recommendation result

Fig. 6(a)-6(e) illustrates Top 5, Top 10, Top 15, Top 20, and Top 25, recommendation results. WAS-PISAR and PISAR have the same start point, because there are none of the user interaction data in the system. Therefore, WAS-PISAR and PISAR cannot perform the filtering process. WTQI give a better results at begin, because WTQI use Wikipedia pre-filter and low level image feature to enhance the tags. WAS-PISAR offer a sharp improvement in recommendation results after 3000 transactions because of the user interaction data. WTQI do not generate better results after performing the similarity calculation. WTQI adopts the process to the tags once, therefore, WTQI gets a horizontal line.

The recommendation results have apparently improved with WAS algorithm. Top 5 increases 13.7%, Top 10 increases 21%, Top 15 increases 24.42%, Top 20 increases 31.94%, and Top 25 increases approximately 40% in accuracy when compared with PISAR system, as shown in Fig. 6(f).

E. Image Recommendation Results

As shown in Fig. 7, the image recommendation results which adopt with WAS algorithm have better relevance than

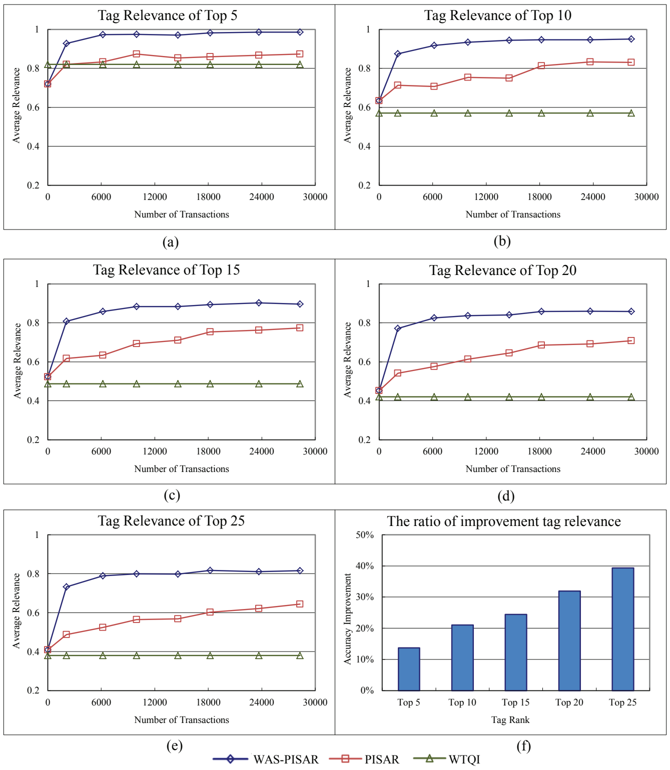


Fig. 6. The tag recommendation results.

PISAR and WTQI. The tag recommendation results achieve more significant growth than do the tag recommendation results. The major reason is that the relationship among tags are the instincts for reviewers. Therefore, reviewers can decide the comments directly. However, the images contain lot of information than the tags do, such as visual objects, concepts, colors or textures. These elements influence their opinion when they are commenting. For example, the image of the sky and the image of the sea may cause reviewers to think they are analogous because they are both blue. Nevertheless, WAS-PISAR still possesses better results for the image recommendation.

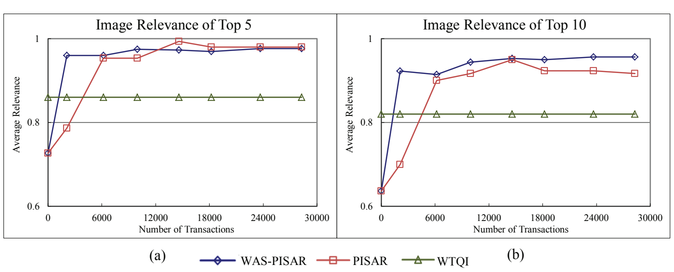


Fig. 7. The image recommendation results.

V. Conclusion

With the rapid growth of social tagging Web sites, tags with bad qualify such as the irrelevant information or typos reduced the accuracy of the recommendation results and wasted the computational resources. We proposed WAS algorithm to overcome the obstacle of bad-quality of tags by collecting users' interaction records. WAS algorithm included two modules, the importance filter and the relationship retriever, to assign each tags into either waking state or sleeping state. These tags which were assigned to the waking state having high contribution to images; otherwise, they were designated as sleeping tags. The tags in waking state improved the quality of recommendation result and reduced the computational resources. The experiments proved that our idea was feasible, and the accuracy of the Top 25 tag recommendation results were about 40% higher than using PISAR system. The image and tag recommendation results were progressively improved with WAS algorithm.

References

[1] B. Sigurbjörnsson and R. van Zwol, "Flickr tag recommendation based on collective knowledge," in *Proceedings of the 17th international conference on World Wide Web*, 2008, pp. 327–336.

[2] L. Wu, L. Yang, N. Yu, and X.-S. Hua, "Learning to tag," in *Proceedings of the 18th international conference on World wide web*, 2009, pp. 361–370.

[3] D. Liu, M. Wang, L. Yang, X.-S. Hua, and H. Zhang, "Tag quality improvement for social images," in *Proceedings of the 2009 IEEE international conference on Multimedia and Expo*, 2009, pp. 350–353.

[4] J.-W. Huang, C.-Y. Tseng, M.-C. Chen, and M.-S. Chen, "Pisar: Progressive image search and recommendation system by auto-interpretation and user behavior," in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, 2011, pp. 1442–1447.

[5] M. Ames and M. Naaman, "Why we tag: motivations for annotation in mobile and online media," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2007, pp. 971–980.

[6] L. Xie, A. Natsev, M. Hill, J. R. Smith, and A. Phillips, "The accuracy and value of machine-generated image tags: design and user evaluation of an end-to-end image tagging system," in *Proceedings of the ACM International Conference on Image and Video Retrieval*, 2010, pp. 58–65.

[7] P. Wu, S. C.-H. Hoi, P. Zhao, and Y. He, "Mining social images with distance metric learning for automated image tagging," in *Proceedings of the fourth ACM international conference on Web search and data mining*, 2011, pp. 197–206.

[8] S. Gao, Z. Wang, L.-T. Chia, and I. W.-H. Tsang, "Automatic image tagging via category label and web data," in *Proceedings of the international conference on Multimedia*, 2010, pp. 1115–1118.

[9] M. Zhou, S. Bao, X. Wu, and Y. Yu, "An unsupervised model for exploring hierarchical semantics from social annotations," in *Proceedings of the 6th international The semantic web and 2nd Asian conference on Asian semantic web conference*, 2007, pp. 680–693.

[10] X. Wu, L. Zhang, and Y. Yu, "Exploring social annotations for the semantic web," in *Proceedings of the 15th international conference on World Wide Web*, 2006, pp. 417–426.

[11] Z. Xu, Y. Fu, J. Mao, and D. Su, "Towards the semantic web: Collaborative tag suggestions," in *Proceedings of the Collaborative Web Tagging Workshop at the World Wide Web*, 2006.

[12] F. Carmagnola, F. Cena, O. Cortassa, C. Gena, and I. Torre, "Towards a tag-based user model: How can user model benefit from tags?" in *Proceedings of the 11th International Conference on User Modeling*, 2007, pp. 445–449.

[13] D. Laniado, D. Eynard, and C. M. olombetti, "Using wordnet to turn a folksonomy into a hierarchy of concepts," in *Semantic Web Application and Perspectives - Fourth Italian Semantic Web Workshop*, 2007, pp. 192–201.

[14] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Information Processing and Management*, vol. 24, no. 5, pp. 513–523, 1988.