International Journal of Scientific Engineering and Technology Research
www.ijsetr.com
ISSN 2319-8885
Vol.03,Issue.31
October-2014,
Pages:6126-6133

To Strengthen the Similarity Integration in Heterogeneous Image-Rich Information Networks

MOHAMMED MOHIULLAH KHAN[1], ABDUL RASOOL MD[2], ABDUL MAJEED[3], D. SRIKANTH[4]
[1]PG Scholar, Dept of CSE, LORDS Engineering College, Hyderabad, India, E-mail: mohammedmohiullahkhan@gmail.com.
[2]Assoc Prof, Dept of CSE, LORDS Engineering College, Hyderabad, India, E-mail: rasool.501@gmail.com.
[3]HOD, Dept of CSE, LORDS Engineering College, Hyderabad, India, E-mail: mjd4001@gmail.com.
[4]Assoc Prof, Dept of CSE, LORDS Engineering College, Hyderabad, India, E-mail: sreekanth.543d@gmail.com.

Abstract: Social multimedia sharing and hosting websites, such as Flickr and Face book, contain billions of user-submitted images. Popular Internet commerce websites such as Amazon.com are also furnished with tremendous amounts of product-related images. In addition, images in such social networks are also accompanied by annotations, comments, and other information, thus forming heterogeneous image-rich information networks. In this paper, we introduce the concept of (heterogeneous) image-rich information network and the problem of how to perform information retrieval and recommendation in such networks. We propose a fast algorithm heterogeneous minimum order *k*-Sim Rank (HMok-Sim Rank) to compute link-based similarity in weighted heterogeneous information networks. Then, we propose an algorithm Integrated Weighted Similarity Learning (IWSL) to account for both link-based and content based similarities by considering the network structure and mutually reinforcing link similarity and feature weight learning. Both local and global feature learning methods are designed. Experimental results on Flickr and Amazon data sets show that our approach is significantly better than traditional methods in terms of both relevance and speed. A new product search and recommendation system for e-commerce has been implemented based on our algorithm.

Keywords: Information Retrieval, Image Mining, Information Network, And Ranking.

I. INTRODUCTION

Social multimedia (photo and video) sharing and hosting websites, such as Flickr, Face book, YouTube, Picasa, Image Shack, and Photo bucket, are popular around the world, with over billions of photos uploaded by users. Popular Internet commerce websites such as Amazon are also furnished with tremendous amounts of product-related images. In addition, many images in such social networks are accompanied by information such as owner, consumer, producer, annotations, and comments. They can be modeled as heterogeneous image-rich information networks. Fig.1 shows an example of the Flickr information network, where images are tagged by the users and image owners contribute images to topic groups. Fig.2 shows an Amazon information network of product images, categories, and consumer tags. Conducting information retrieval in such large image rich information networks is a very useful but also very challenging task, because there exists a lot of information such as text, image feature, user, group, and most importantly the network structure.

In text-based retrieval, estimating the similarity of the words in the context is useful for returning more relevant images. Word Net manually groups words into synonym sets; Google Distance [3] computes word similarity by co-occurrence in search results. Flickr Distance [4] considers visual relationship. In image content-based retrieval, most methods (such as Google's Visual Rank [4]) and systems [6], [7], [8], [9], [10] compute image similarity based on image content features. Hybrid approach combines text features and image content features together [11], [12]. Most commercial image search engines use textual similarity to return semantically relevant images and then use visual similarity to search for visually relevant images. Integration-based approaches [12] use linear or nonlinear combination of the textual and visual features. However, existing works cannot handle the link structure. In this paper, we propose an image-rich information network model where the similarities between same type of nodes and different types of nodes can be better estimated based on the mutual impact under the network structure.

Among algorithms that compute object similarity in information networks, Sim Rank is one of the most popular, but it is very expensive to calculate and the similarity is only based on the link information. When consider the images in the network, image similarity can actually also be judged by content features, such as RGB histogram and SIFT. In this paper, we propose an efficient approach called MoK-Sim Rank to significantly improve the speed of Sim-Rank, and introduce its extension HMok-Sim Rank to work on weighted heterogeneous information networks. Then, we propose algorithm IWSL to provide a novel way of integrating both link and content information. IWSL performs content and link reinforcement style learning with either global or local feature weight learning.

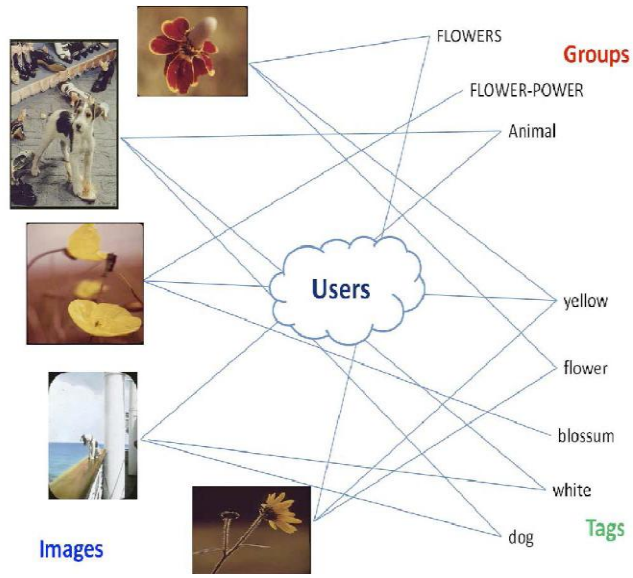

Fig.1. Information network for Flickr, connected by images, user tags, and groups

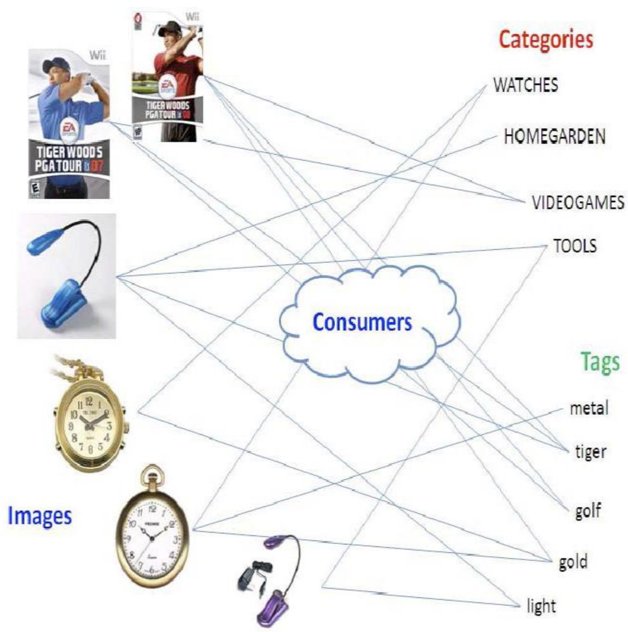

Fig.2.Information network for Amazon, connected by products, user tags, and categories.

Based on the proposed algorithm, a novel product recommendation system has been implemented for ecommerce to find both visually and semantically relevant products modeled in an image-rich information network. Fig. 3 describes the system architecture. The bottom layer contains the product data warehouse which includes product images and related product information. The second layer performs meta information extraction and image feature extraction. The third layer builds a weighted heterogeneous image-rich information network. The fourth layer performs information network analysis based ranking to find relevant results for a query. The top layer contains a user-friendly interface, which interacts with users, responds to their requests, and collects feedback. The remainder of the paper is organized as follows: SectionII describes System Design. SectionII reports experimental results. SectionIV concludes the paper.

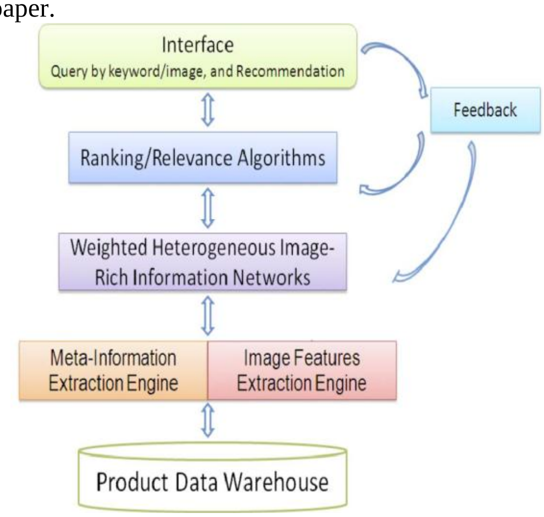

Fig.3.Product search and recommendation system architecture.

II. SYSTEM DESIGN

A. Architecture

The Mok-Sim Rank algorithm measures the link based similarity by finding the similarity of group and similarity of tags separately then adding them. Cosine similarity is used to calculate the content similarity between the two images Finally apply integration technique to combine the link based and content based similarity measure then classifies the images using classifier. These measures are used by our recommendation system to suggest social image resource to the users.

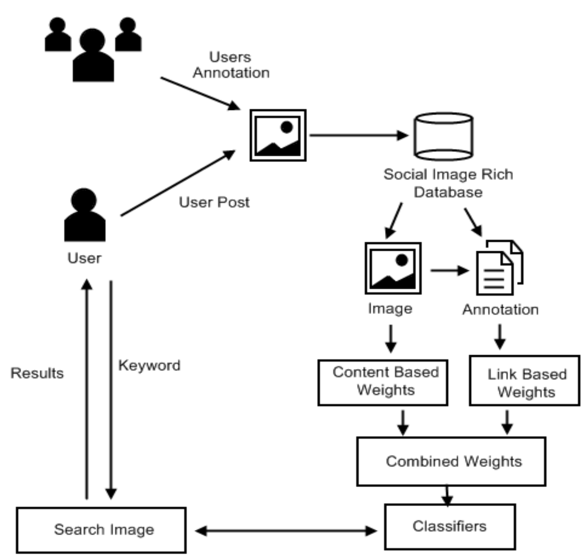

Fig.4. System Architecture.

B. Modules

1. Social Community Development

This module is used to construct the basic Social multimedia platform similar to the flicker through which the image and image related data is collected and preprocesses step to construct the required data model for our Similarity Integration. This module contains user registration and login process, image posting, sharing and human based annotation like tagging of images and image preprocessing step.

2. Link-Based Similarity

Similar images are likely to link to similar tags and groups, so we define the link-based semantic similarity between images as combination of similarity of group and similarity of tags. It is defined as follows

$$S_{m+1}(e,e') = \alpha_1 S_m^G(e,e') + \beta_1 S_m^T(e,e') \qquad (1)$$

This module iteratively calculate the similarity between image pairs, similarity between group pairs of images and similarity between tag pairs of image until the convergence is reached.

3. Content-Based Image Similarity

The image vector information is extracted from the image content based on color and histogram and this vector information is used by the cosine similarity function to measure the similarity. Cosine similarity is a measure of similarity between two vectors of an inner product space that measures the cosine of the angle between them. The cosine of 0° is 1, and it is less than 1 for any other angle. It is thus a judgment of orientation and not magnitude: two vectors with the same orientation have a Cosine similarity of 1, two vectors at 90° have a similarity of 0, and two vectors diametrically opposed have a similarity of -1, independent of their magnitude.

4. Similarities Integration

We present novel algorithm to integrate link-based and content-based similarities: First perform HMok-Sim Rank to compute the link-based similarities and second perform feature learning considering the link-based similarity to update the feature weights, and then update the node similarities based on the new content similarity.

5. Formation of Clustering

The final weight calculated after integration of both link based similarity and content similarity is used to classify the images based on their similarity. From this clustered information users can retrieve their needed images by using keyword.

III. EXPERIMENTS

A. Data Sets

We conduct experiments on two data sets: Flickr and Amazon. The Flickr data set is created by downloading the images and related metadata information, such as groups and tags using Flickr API. The Amazon data set is created by downloading product images and related metadata information, such as category, tags, and title, via the API of Amazon. The Amazon API only returns the top five tags for each product, so we use the words in the title as additional tags. Product category is treated as group. Table 1 shows the statistics for the two data sets. For image feature extraction, we extracted CEDD, which is a compact descriptor that considers both color and edge features. In literature, it has shown good performance compared with many traditional features. Note that our model is general to other features or combination of them. Tag preprocessing. We change all tags to lower case. Tags with only number characters are removed. Stop-words, such as the, you and me, are also removed. The remaining tags are stemmed using the Porter stemming algorithm.

Fig.5. Tag frequency for the Flickr data set *Y* -axis denotes the frequency, *X*-axis denotes the ordered (by frequency) tag id.

We remove very infrequent tags and only retain those that appear in more than k (e.g., $k = 2$) images. Fig.5 shows the tag frequency (the number of images annotated with the tag) of data set Flickr. We can observe that many tags have low frequency, and a very few tags have high frequency. The Amazon data set has similar curve shape. Table 1 shows the top 10 most frequent tags for the two data sets.

B. Experimental Setting

Experiments were conducted on a PC with Intel Pentium(R) D 3.4 GHz CPU and 4 GB RAM, running Windows XP.

C. Speed Performance

Fig.6 shows the speed-up ($T_{baseline}/T_i$) of HK-Sim Rank and HMoK-Sim Rank over baseline H-Sim Rank. With the increase of the number of images, they become increasingly faster than H-Sim-Rank. Note that we only show result for as most 3,000 images due to the expensive time complexity of the baseline H-Sim Rank, which takes too long for larger data. Fig. 7 shows the speed-up of HMoK-Sim Rank over HK-Sim Rank. We can see that HMoK-Sim Rank is much faster than HK-Sim Rank. Because IWSL is based on HMoK Sim Rank, it has similar time efficiency except for the time spent on feature weight learning. For exact executing time, we take the Amazon data set as an example: when the number of images is 3,000, H Sim Rank takes 598 seconds, *HK*-Sim Rank takes 24 seconds, and HMoK-Sim Rank takes 9 seconds; when the number of image increase to 20,000, *HK*-Sim Rank takes 928 seconds, while HMoK-Sim Rank only takes 132 seconds.

TABLE I: Top 10 Most Frequent Tags

(a) Flickr

Tag	Frequency
flower	1970
nikon	1545
night	1491
white	1468
black	1288
canon	1252
geotag	1250
light	1231
sky	1127
nature	1122

(b) Amazon

Tag	Frequency
black	6088
pack	4621
case	4235
watch	3987
women	3706
men	3611
classic	3348
music	3321
set	3231
game	3228

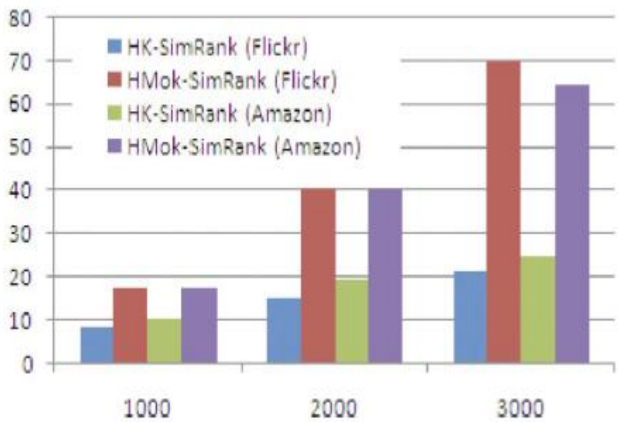

Fig.6. Speed-up of *HK*-Sim Rank and HMok-Sim Rank over H Sim Rank *X*-axis denotes the number of images, *Y* -axis denotes the speed-up (in times ratio).

D. Relevance Performance

1. General Result

Because of the large number of images, it is difficult to check one by one to obtain a complete set of relevant images for each query image. In order to generate an approximate ground truth for performance evaluation, we assume that if two images are relevant, their visual similarity should be above a threshold εv and the number of shared tags should also be above a threshold εt. We ignore images which contain less than five tags. Such images make up 6.5 and 15.3 percent of the Flickr and Amazon data set, respectively. Since all considered images have more than five tags, which are human annotations, if two images don't have any common tag, it is likely that the users who made those tags do not think they are relevant. On the Internet, there are many images do not have tags, to simulate the real world case; we randomly select 50 percent images to remove all their tags. Our algorithm can still find some of them as relevant because we are able to learn a feature weight based on those images which have tags or other link-based information.

We compare VLWC (weighted combination of visual and link similarity without feature weight learning), IWSL_L (IWSL with local feature weight learning), and IWSL_G (IWSL with global feature weight learning) to several baselines: Visual (only use the visual similarity), Text (only use the textual similarity, following a popular text retrieval approach: cosine measure based on the *tf* * *idf* weighted tag vector), VTWC [12] (weighted combination of visual and textual similarity, we choose equal weight), Link (HMok-Sim Rank which only use the link similarity), Min Fusion and Max Fusion [12].

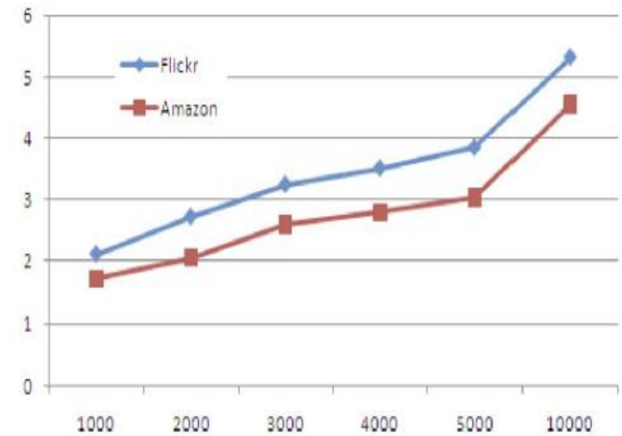

Fig.7. Speed-up of HMok-Sim Rank over HK-Sim Rank *X*-axis denotes the number of images, *Y* -axis denotes the speed-up (in times ratio).

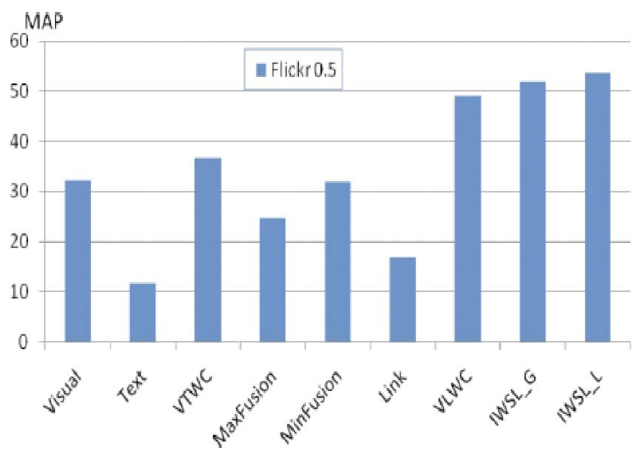

Fig.8. MAP of the algorithms on Flickr data. *X*-axis denotes the algorithms. *Y* -axis denotes the MAP (percent).

Evaluation method we use mean average precision (MAP) to measure the retrieval performance of the algorithms. For every image in the data set, we obtain a ranking list of relevant images computed by each algorithm and compute the average precision based on the approximate ground truth before removing tags. The final MAP score for each algorithm is calculated as the mean average precision of each image. Note that there is no training data so all the algorithms are unsupervised. Figs.8 and 9 show the result on Flickr and Amazon data, respectively. We can see that link-based similarity performs better than text-based similarity; VLWC achieves better performance than traditional algorithms by linearly combining visual and link information together. Algorithm IWSL further improves the performance by introducing a new way of integrating content and link information via mutual reinforcement with feature learning. IWSL_L achieves better results than IWSL_G, because IWSL_L performs local feature learning, which can find a specific and better feature weighting for each image than global feature learning, which finds a general feature weighting for all images.

2. Case Study

As an example from the Flickr data set, Fig.10 shows the top 10 most similar images for a query image about "moon," using link-based (Sim Rank) (first row), content based similarity (second row), and IWSL (third row), respectively. The top left image is the query image. Clearly, IWSL obtains the most relevant matches for both semantic and visual appearances. In another example from the Amazon data set, Fig.11 shows the top 10 most similar images for a query image.

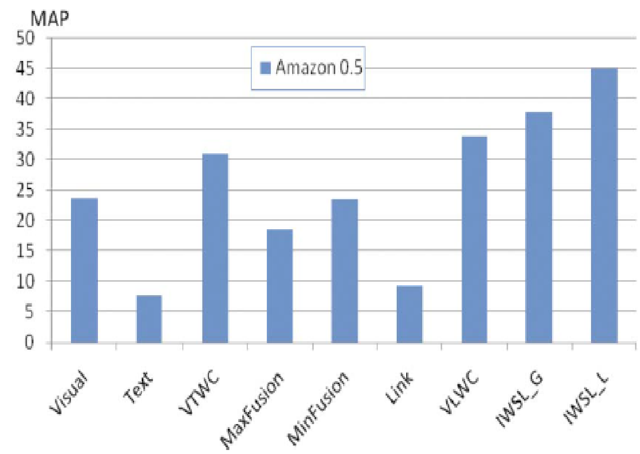

Fig.9. MAP of the algorithms on Amazon data. X-axis denotes the algorithms. *Y* -axis denotes the MAP (percent).

about "i-Phone," using link-based similarity (Sim Rank) (first row), content-based similarity (second row), and IWSL (third row). Again, IWSL obtains the best results in terms of the relevance for both semantic and visual appearances. Our experiments also show good performance of our algorithm to find similar groups and relevant tags. For example, similar groups about night are "Night Images," "No-Flash Night Shots," "After Dark - Night Photography" and "Night Lights." Relevant tags to flower are "floral," "flora," and "botani." We can use such tag similarity to help find more relevant images for a keyword query.

E. Parameter Setting

The experiments are based on the following parameter setting: τ = 0.5, $\mu = F_{ij}$ and, β = 0.5 and all γ parameters as 0.5 for Gradient Decent. The damping factor parameters are set as 0.8 by following the standard of Sim Rank. As an example to demonstrate how we obtain the optimal parameter setting, Fig.12 shows the MAP of algorithm IWSL_L w.r.t. parameter β, for both data sets. When β is too small which means significantly ignoring the regulator, the performance is not good, which proves the benefit of introducing the regulator; when β is too big, the performance is also not good, because the regulator will dominate the optimization.

F. Convergence

Figs.13 and 14 show $\Delta S = |S_{m+1} - S_m|$, the absolute change in the average sum of similarity scores (including images, groups, and tags) from iteration m to m+1 for algorithm IWSL_L on data set Flickr and Amazon, respectively.

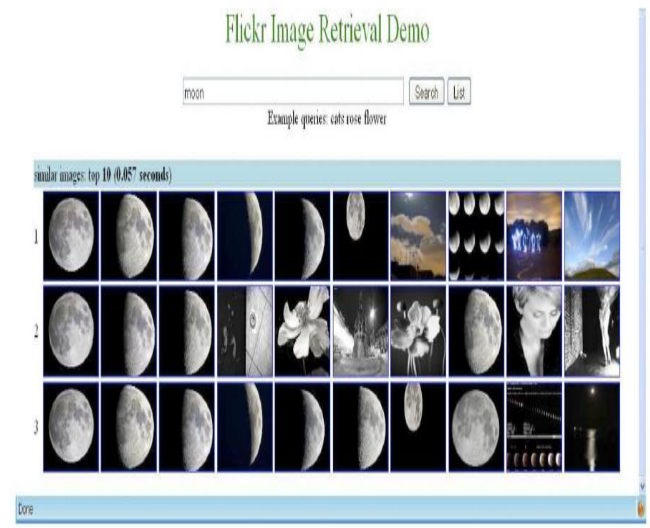

Fig.10. Top 10 retrieval results by Link (Sim Rank), Content similarity, and IWSL the top left image is the query image from Flickr. It is tagged with "moon, lune, sky" and belongs to group "After Dark - Night Photography."

Fig.11. Top 10 retrieval results by Link (Sim Rank), Content similarity, and IWSL the top left image is the query image from Amazon. It is tagged with "iPhone, invisible shield, accessories" and belongs to category "Wireless Accessories."

IWSL_G has similar results. We can see that IWSL converges very fast, and only five to six iterations are enough for most scenarios.

F. Product Search and Recommendation System for E-Commerce

Fig.15 shows our product search and recommendation system for e-commerce using Amazon products as an example when users search and click on a product in a web browser, we can recommend both visually and semantically relevant products, with the relevance score computed by our IWSL algorithm. Fig.16 shows a comparison of our recommendation with the Amazon recommendation based on “Customers Who Bought This Item Also Bought” (which we call cobought). When a consumer wants to buy a bag, our approach provides more relevant recommendations for the product. Fig.17 shows another comparison with the Amazon recommendation based on “Customers Who Bought This Item Also Viewed” (which we call coviewed). The results are similar, because when a Amazon user wants to buy a jewelry, he or she is likely to browse through similar products before making a final choice.

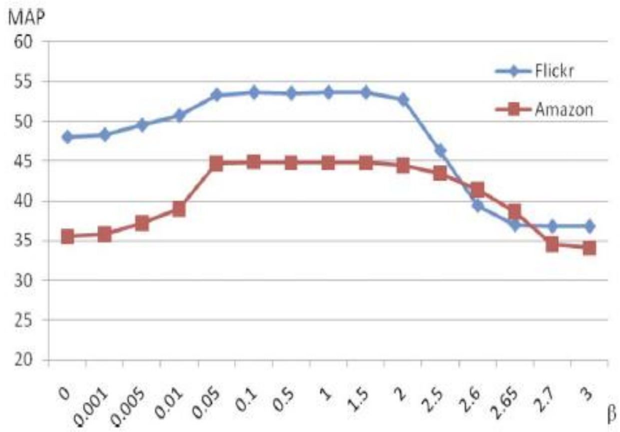

Fig.12. MAP w.r.t. parameter *β*. *X*-axis denotes *β*. *Y* -axis denotes MAP.

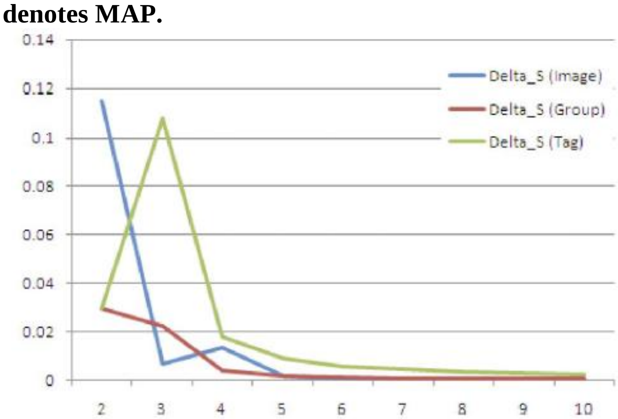

Fig.13. Convergence of IWSL on Flickr data *X*-axis denotes the number of iteration. *Y* -axis denotes ∆S (i.e., *Delta _S*).

One problem of using coviewed or cobought information for recommendation is that only the top-k ranking list is available, but the details (e.g., the frequency of such co-occurrence) are commercial secret and are not publicly accessible. Without such proprietary information, it is difficult to combine sources from multiple e-commerce websites, such as Amazon, Best Buy, Wal Mart, and Target, to generate an overall recommendation. We believe this is one of the reasons why general product search engines, such as Google Product Search and Bing Shopping, currently do not have such recommendation functions.

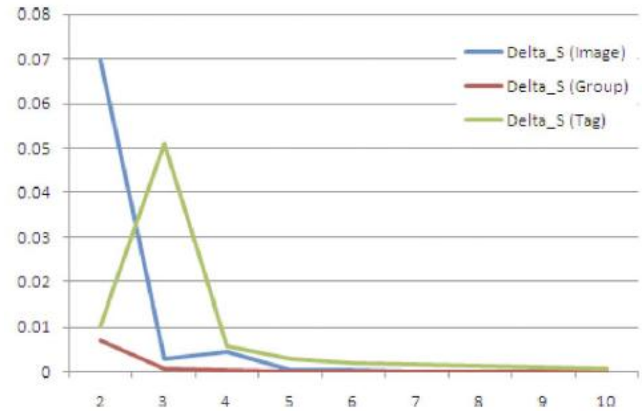

Fig.14. Convergence of IWSL on Amazon data *X*-axis denotes the number of iteration. *Y*-axis denotes ∆S (i.e., *Delta _S*).

Fig.15. Snapshot of the product search and recommend-ation system for e-commerce.

However, for products (e.g., jewelry, watch, bag, glasses, and clothes) that depend heavily on visual appearance to attract consumers, we are able to generate both visually and semantically relevant recommendations without the coviewed or cobought information. This leads to a better general product search service that enables users to compare and make the best choice. Similar strategy can be applied to Flickr for photo and interest recommendation. When users browse Flickr photos, we can recommend relevant Flickr photos, or interest groups for users to join. In addition, by integrating Flickr and Amazon networks, we can recommend relevant Amazon product photos to a Flickr photo for advertisement.

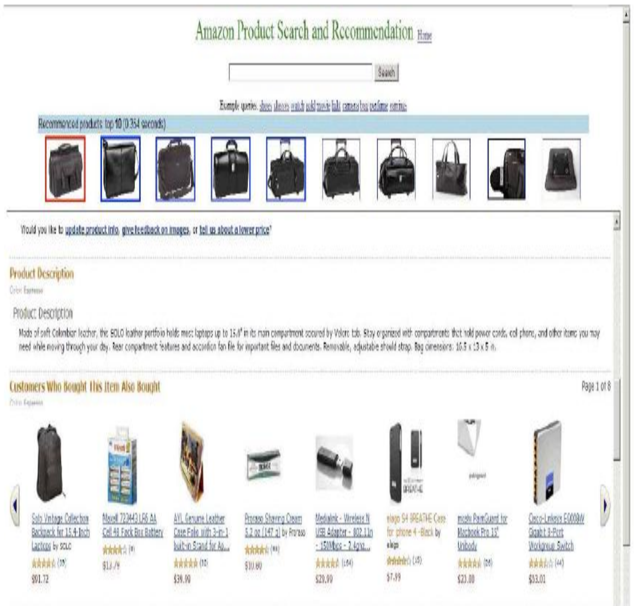

Fig.16. Recommendation comparison, ours versus Amazon's "Customers Who Bought This Item Also Bought"

Fig.17. Recommendation comparison, ours versus Amazon's "Customers Who Bought This Item Also Viewed"

IV. CONCLUSIONS AND FUTURE WORK

This paper presents a novel and efficient way of finding similar objects (such as photos and products) by modeling major social sharing and e-commerce websites as image rich information networks. Our major contributions are as follows:

1. We propose HMok-Sim Rank to efficiently compute weighted link-based similarity in weighted heterogeneous image-rich information networks. The method is much faster than heterogeneous Sim Rank and K-Sim Rank.
2. We propose both global and local feature learning approaches for learning a weighting vector to capture more important feature subspace to narrow the semantic gap.
3. We propose the algorithm IWSL to provide a novel way of reinforcement style integrating with feature weighting learning for similarity/relevance computation in weighted heterogeneous image-rich information network.
4. We conduct experiments on Flickr and Amazon networks. The results have shown that our algorithm achieves better performance than traditional approaches.
5. We have implemented a new product search and recommendation system to find both visually similar and semantically relevant products based on our algorithm.

Future work Under the concept of heterogeneous image rich information network, many future works are in our sight. It will be interesting to see how such kind of network structure may benefit various image mining and computer vision tasks, such as image categorization, image segmentation, tag annotation, and collaborative filtering. As for the proposed algorithm IWSL, we plan to study the problem of how to get an optimal combination of both local and global learning to achieve a balance on time and quality performance. In order to use IWSL in web scale search engine, a distributive computing extension will be investigated. Considering dynamic environment is also important. One potential solution could be: First, perform network clustering to partition the whole large network into small connected components as sub networks; Second, run the proposed algorithm on each sub network; and third, when a new image come, update the sub network which the new image belongs to.

V. REFERENCES

[1] Xin Jin, Jiebo Luo, Fellow, IEEE, Jie Yu, Member, IEEE, Gang Wang, Member, IEEE, Dhiraj Joshi, Member, IEEE, and Jiawei Han, Fellow, IEEE, "Reinforced Similarity Integration in Image-Rich Information Networks", IEEE Transactions on Knowledge and Data Engineering, Vol. 25, No. 2, February 2013.

[2] X. Jin, J. Luo, J. Yu, G. Wang, D. Joshi, and J. Han, "iRIN: Image Retrieval in Image-Rich Information Networks," Proc. 19th Int'l Conf. World Wide Web (WWW '10), pp. 1261-1264, 2010.

[3] R.L. Cilibrasi and P.M.B. Vitanyi, "The Google Similarity Distance," IEEE Trans. Knowledge and Data Eng., vol. 19, no. 3, pp. 370-383, Mar. 2007.

[4] L. Wu, X.-S. Hua, N. Yu, W.-Y. Ma, and S. Li, "Flickr Distance," Proc. 16th ACM Int'L conf. Multimedia, pp. 31-40, 2008.

[5] Y. Jing and S. Baluja, "Visual Rank: Applying Page rank to Large- Scale Image Search," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 30, no. 11, pp. 1877-1890, Nov. 2008.

[6] R.C. Veltkamp and M. Tanase, “Content-Based Image Retrieval Systems: A Survey,” technical report, Dept. of Computing Science, Utrecht Univ., 2002.
[7] H. Tamura and N. Yokoya, “Image Database Systems: A Survey,” Pattern Recognition, vol. 17, no. 1, pp. 29-43, 1984.
[8] W.I. Grosky, “Multimedia Information Systems,” IEEE Multi-Media, vol. 1, no. 1, pp. 12-24, Spring, 1994.
[9] V.N. Gudivada and V.V. Raghavan, “Content-Based Image Retrieval Systems,” Computer, vol. 28, no. 9, pp. 18-22, Sept. 1995.
[10] Y. Rui, T.S. Huang, and S.-F. Chang, “Image Retrieval: Current Techniques, Promising Directions, and Open Issues,” J. Visual Comm. and Image Representation, vol. 10, no. 1, pp. 39-62, 1999.
[11] R. Datta, D. Joshi, J. Li, and J.Z. Wang, “Image Retrieval: Ideas, Influences, and Trends of the New Age,” ACM Computing Surveys, vol. 40, no. 2, pp. 1-60, Apr. 2008.
[12] T. Deselaers and H. Mller, “Combining Textual- and Content-Based Image Retrieval, Tutorial,” Proc. 19th Int’l Conf. Pattern Recognition (ICPR ’08), http://thomas .deselaers.de/ teaching/ files/tutorial_icpr08/04-combination.pdf, 2008.