International Journal of Emerging Trend in Engineering and Basic Sciences (IJEEBS)
ISSN (Online) 2349-6967 *Volume 2 , Issue 1(Jan-Feb 2015), PP1-5*

Reinforced Similarity Integration In Image-Rich Information Networks

Kaustubh Bhattad[1], Amol Gore[2], Rohit Jadhav[3], Raghunandan Joshi[4]
Computer Engineering, JSPM`s Rajarshi Shahu College of Engineering/Pune University, India

Abstract :-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 image-rich information network and the problem of how to perform information retrieval and recommendation in such 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, ranking

I. INTRODUCTION

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 image-rich information networks. In this paper, we introduce the concept of 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 SimRank 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. It gives us more refined and filtered image search rather than existing one by using various algorithms applied on it. It also comes with text based as well as image based retrieval of images from existing data set.

II. EXISTING SYSTEM

Image Retrieval System is a computer system for Browsing, Searching and retrieving image from large database. The use of metadata such as captioning, keywords or decryptions to the images store in the database along with the images or low level features extracted from the image like shape, color and texture have been use till now for the image retrieval from existing search engine.

SimRank store text in the form of keywords together with the image. SimRank uses surrounding text to search the keywords which are physically close to the image. This technique realize on the assumptions that the surrounding text describe the image. Search engines that use that technique are Google, yahoo and ulta vista.

In this technique each region as a separate object and then check the similarity between object. It can correctly separate the regions that have same properties that we define. Region growing methods can provide the original images which have clear edges and good segmentation results. The existing system is effective one but not much.

III. PROPOSED SYSTEM

We propose image search engine, an Applicationaware refinement of images from data set scheme that not only exploits application awareness, but also combines result of SimRank and CBIR, to achieve high filtration efficiency of images by refining the large data set latency to limited number of image set which are more closer to query and feedback image. Our application-aware design is motivated by efficient use of SimRank and CBIR combinely.We observe that there is a significant difference among different types of applications in the personal computing environment in terms of data redundancy, sensitivity to different refinement methods.

Thus, the basic idea of image filtering is to effectively exploit this application difference and awareness by treating different types of applications independently and adaptively during the Simrank and CBIR processes to significantly improve the refinement efficiency and reduce the system overhead. Plus as part of adding something new to it we propose various algorithm for feature extraction and edge detection. The aim and objective of our paper is that to employ an intelligent Filtration, adaptive use text and content based retrieval on application awareness and to create application-specific image search engine.

IV. ARCITECTURE OVERVIEW

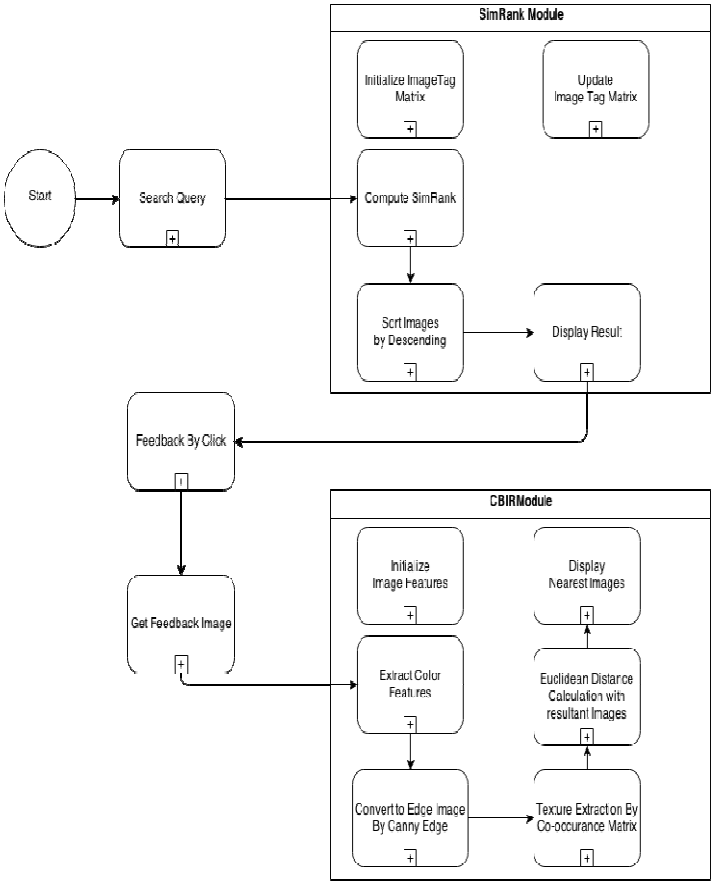

Fig. 1 An architectural overview of retrieving image

An architectural overview of retrieving image is illustrated in fig 1. where initialization of image matrix is pre-defined by taking reference of existing data set. User first search the query as an input. Input query is nothing but the name for image that he/she wants to search from data set which must be filtered. For making the search more efficient architecture is divided into two vital modules.

a. **SimRank module**

It consists various blocks for retrieving images by text which already contains predefined matrix of image tag. This block applies simRank algorithm to the given query. We get finite set of result as the output of simrank. Output will set in such a fashion image having more valu computed by simRank should display first. This filtered set of images display to user as an output of simRank algorithm. One more bock is supported by SimRank which is feedback. By clicking on one image it will give refined result again which relates to that image.

b. **CBIR module**

This block takes input as feedback image and works on its content as it is the content based image retrieval search. Image features are defined already as they are the part of initialization. CBIR is mainly focuses on various image detection algorithms which makes it more efficient and can easily work under any circumstances. The image which user preferred as a feedback image CBIR module performs feature extraction on it for further refinement. At every phase we are getting more filtered set of images than previous one. Conversion to edge image is also done within CBIR module. In some cases only feature extraction is not inadequate. Image is converted in the edge image by using canny edge algorithm.

Again to make our result more refined texture extraction is done on the given set of images. Texture extraction is done by using co-occurrence matrix in efficient way. It will shows how many times it is occurred. Euclidean distance calculation for calculating ordinary distance between two points or pixel present in the image. After calculating it display the limited set of images as they are the nearest images to feedback image.

V. CONCLUSION

We propose HMok-SimRank to efficiently compute weighted link-based similarity in weighted hetero- geneous image-rich information networks. The method is much faster than heterogeneous SimRank and K-SimRank. 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. 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.

VI. FUTURE WORK

Under the concept of heterogeneous image rich information network, many future works are in our sight. It will 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.

REFERENCES

[1]. Z. Yang and C.-C.J.Kuo, *"Survey on Image Content Analysis, Indexing, and Retrieval Techniques and Status Report of Mpeg-7,"* Tamkang J. Science and Eng., vol. 3, no. 2, pp. 101-118, 1999.

[2]. J. Huang,S.R. Kumar, M. Mitra, W.-J. Zhu, and R. Zabih, *"Image Indexing Using Color Correlograms,"* Proc. IEEE CS Conf. Computer Vision and Pattern Recognition (CVPR '97), pp. 762-768, 1997.

[3]. S. Chatzichristofis and Y. Boutalis*, "CEDD: Color and Edge Directivity Descriptor: A Compact Descriptor for Image Indexing and Retrieval,"* Proc. Sixth Int'l Conf. Computer Vision Systems, pp. 312-322, 2008.

[4]. S. Aksoy and R.M. Haralick, *"Textural Features for Image Data base Retrieval,"* Proc. IEEE Workshop Content - Based Access of Image and Video Libraries (CBAIVL '98), p. 45, 1998.