ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 2, Febraury 2014

Image Retrieval Based on Content Search Mechanism

R.Keerthika[1], Dr.C.Nalini[2]

Assistant Professor, Department of Information Technology, Karpagam College of Engineering, Coimbatore, India[1]

Professor, Department of Information Technology, Kongu Engineering College, Perundurai, India[2]

ABSTRACT: Searching a image from search engines and social networks are providing results based on the comments, clarification and some other information about the images. The results will be generated easily, Due to ineffective similarity mining the systems may suffer from the accuracy. This paper tells how to overcome the problem of image mining based on effective meta informations and semantic similarity measures. The semantic similarity contains both textual and visual similarity measures. To resolve the mentioned problems, Content-based image retrieval (CBIR) avoids the use of textual descriptions and instead retrieves images based on the content similarities like colours, shapes, textures etc to the user-specified image features. We proposed a method named Content-Based Relevance Response (CBRR) is used to achieve the high retrieval quality by using the discovered patterns. In case of efficiency, the patterns are mined from the user query log can be viewed as the shortest paths to the user space. According to the patterns, the users can obtain a set of relevant images in an online query refinement process. Similarity for the search can be done by using meta tags, shape/region attributes, and colour distribution in images.

KEYWORDS: personalized image search, image retrieval, ranking, datamining.

I.INTRODUCTION

Data mining is the process of extraction of hidden information from large databases, and it is a powerful technology with great potential to help companies focus on the most important information in their data warehouses. Its tools predict future trends. It is a tool that can answer business questions that traditionally were too time consuming to resolve our databases for hidden patterns, because it lies outside their beliefs. Data mining techniques can be implemented rapidly on existing software and hardware platforms to enhance the value of existing information resources, and can be integrated with new products and systems as they are brought on-line.

When implemented on high performance client/server or parallel processing computers, data mining tools can analyze massive databases to deliver answers to questions such as, provides an introduction to the basic technologies of data mining.

Data mining techniques are the result of a long process of research and product development. This evolution began when business data was first stored on computers, continued with improvements in data access, and more recently, generated technologies that allow users to navigate through their data in real time.

Data mining is ready for application in the business community because it is supported by three technologies that are now sufficiently mature:

- Massive data collection
- Powerful multiprocessor computers
- Data mining algorithms

From the users point of view, the four steps were revolutionary because they allowed new business questions to be answered accurately and quickly.

The core components of data mining technology have been under development, in research areas such as artificial intelligence, machine learning statistics, Nowadays, these techniques coupled with high-performance relational database engines and data integration, make these technologies are useful for current data warehouse environments.

A) The Scope of Data Mining

Data mining derives its name from the similarities between searching for valuable business information in a large database. Datamining technology can generate new business opportunities by providing these capabilities:

- Automated prediction of trends and behaviours. Data mining automates the process of finding predictive information in large databases.
- Questions that traditionally required extensive hands-on analysis can now be answered directly from the data quickly. A typical example of a predictive problem is targeted marketing. Data mining uses data on past promotional mailings to identify the targets most likely to maximize return on investment in future mailings. Other predictive problems include forecasting bankruptcy and other forms of default, and identifying segments of a population likely to respond similarly to given events.
- Automated discovery of previously unknown patterns. Data mining tools sweep through databases and identify previously hidden patterns in one step.

Data mining techniques can yield the benefits of automation on existing software and hardware platforms, new systems as existing platforms are upgraded and new products developed. When data mining tools are implemented on high performance parallel processing systems, they can analyze massive databases in minutes. Faster processing means that users can automatically experiment with more models to understand complex data.

Databases can be larger in both depth and breadth:

- More columns. Analysts must often limit the number of variables they examine when doing hands-on analysis due to time constraints. Yet variables that are discarded because they seem unimportant may carry information about unknown patterns. High performance data mining allows users to explore the full depth of a database, without preselecting a subset of variables.
- More rows. Larger samples yield lower estimation errors and variance, and allow users to make inferences about small but important segments of a population.

The commonly used techniques in data mining are:

Artificial neural networks (ANN): Non-linear predictive models that learn through training and resemble biological neural networks in structure.

- Decision trees (DT): Tree-shaped structures that represent sets of decisions. It generates rules for the classification of a dataset. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID).

Genetic algorithms (GA): Optimization techniques that use process such as genetic combination, mutation, and natural selection in a design based on the concepts of evolution.

- Nearest neighbor method (NN): A technique that classifies each record in a dataset based on a combination of the classes of the k records most similar to it in a historical dataset where k * 1 called the k-nearest neighbor technique.
- Rule induction: The extraction of useful if-then rules from data based on statistical significance.

Many of these technologies have been in use for more than a decade in specialized analysis tools that work with relatively small amount of data. These capabilities are now evolving to integrate directly with industry-standard data warehouse and OLAP platforms.

II.PROBLEM IDENTIFICATION

Photo and video sharing by using some social websites, such as Facebook, YouTube, Picasa, Flickr, Photo bucket, and Image Shack are mostly used around the world, most popular websites such as Amazon also provides with remarkable amounts of product based images. Additionally, some images in social networks are accompanied by information such as owner, producer, consumer and comments. They can be modelled as heterogeneous image-rich information networks. In Flickr, the images are tagged by the users and image owners contribute images to topic groups. Similarly in Amazon the categories, product images and consumer also tags. Conducting information retrieval in such large image rich information networks is a very useful but it is also a challenging task, because it consists of lot of information such as user, group, image feature, text, and the network structure. In content-based retrieval, to return the more relevant images the similarity of the words in the context will be useful. Image retrieval is for searching and retrieving images from a large database of digital images. Most common methods are used to utilize metadata such as keywords, or descriptions to the images.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 2, Febraury 2014

In image content-based retrieval, most methods and systems compute image similarity based on image content features. Image meta search is based on associated meta data such as text, keywords etc. Content-based image retrieval (CBIR) avoids the use of textual descriptions and instead retrieves images based on the content similarities like colours, shapes, textures etc to the user-specified image features. In CBIR engines image search is based on visual content such as color, shape/object, and texture etc. A Hybrid approach which combines the text features and image content features together. Most commercial image search engines use visual similarity to search visually relevant images and textual similarity to return semantically relevant images. Integration-based approaches use linear or nonlinear combination of the textual and visual features. However, existing works cannot handle the link structure. In this paper, they proposed an image-rich information network model where the similarities between same type of images and different types of images can be better estimated based on the mutual impact under the network structure. Among algorithms that compute object similarity in information networks, SimRank 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. The proposed work is an efficient approach called MoK-SimRank to significantly improve the speed of Sim-Rank, and introduce its extension HMok-SimRank to work on weighted heterogeneous information networks. Then, we propose algorithm Content-Based Relevance Response (CBRR) to provide a way of integrating both link and content information. It performs content and link reinforcement style learning with either global or local feature weight learning.

Computers are loaded up with lots of information about a variety of situations where an answer is known and then the data mining software on the computer must run through that data and distill the characteristics of the data that should go into the model. This model is built and it can be used in related situations. For e.g., the director of marketing for a telecommunications company and you would like to acquire some new long distance phone customers.

As just randomly go out and mail coupons to the general population just as you could randomly sail the seas looking for drawn fortune. In other case it would achieve the results you desired and of course you have the opportunity to do much better than random you could use your business experience stored in your database to build a model.
As the marketing director you have access to a lot of information about all the customers: such as name, age, sex, history and long distance calling. It is also having a lot of information about your prospective customers: such as name, age, sex, credit history etc. the problem is long distance calling usage of these prospects. So we have to concentrate on those prospects with large amounts of long distance usage. It is accomplished by building a model. Table 1 illustrates the data used for building a model for new customer prospecting in a data warehouse.

Table 1 - Data Mining for Prospecting

Information Type	Customers	Prospects
General information (e.g. demographic data)	Known	Known
Proprietary information (e.g. customer transactions)	Known	Target

This model could then be applied to the prospect data to try to tell something about the information that telecommunications Company does not currently have access. This model with new customers can be targeted. Test marketing is one of the excellent sources of data for performing the kind of modeling. Mining the results of a test market representing a broad but relatively small sample of prospects can provide a foundation for identifying good prospects in the overall market. Table 2 shows another common scenario for building models: predict what is going to happen in the future.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 2, Febraury 2014

Table 2 - Data Mining for Predictions

Information Type	Yesterday	Today	Tomorrow
Static information and current plans (e.g. demographic data, marketing plans)	Known	Known	Known
Dynamic information (e.g. customer transactions)	Known	Known	Target

III. PREVENTION AGAINST PERSONALIZED SEARCH

Increasing development in social sharing websites, like YouTube Flickr, allow users to create, share, and comment. Then the larger scale user generates meta-data not only facilitate users in sharing and organizing the content of multimedia, but it provides useful information to develop media retrieval and management. Modified search serves as an example where the web search experience is improved by generating the returned list according to the modified intent of user search. This paper exploits the social explanations and proposes a novel framework simultaneously considering the user and query relevance to learn to modified image search. The basic principle is to embed the user preferences and query-related search intent into user-specific topic spaces. Since the users original annotation is sparse for topic modeling, to enrich users annotation pool before user specific topic spaces construction.

The proposed framework contains two major components:
1) A Ranking based Multi-correlation Tensor Factorization model is proposed to perform explanation calculation, which is considered as users potential explanations for the images;
2) We introduce user specific Topic modelling to map the query relevance and user preference into the same user specific topic space. For performance assessment, these resources involved with users social activities are employed. Experiments on a large scale Flickr dataset demonstrate the effectiveness of the proposed method.

IV.RESULTS

Image processing extracts the necessary image at an accurate time. This image search mechanism is provided in several modules like admin, user, and non-personalized user. These different modules help us to upload image, delete image and can get an accurate image. User can create account with the username and password. He can add photo, view the photos that are uploaded, he can also tag the photos to the particular friends he wants to. Administrator has the main authority for accessing the user account, can view the user activities like what he has uploaded. Administrator can also delete the unnecessary images if he feels so. Whatever photo user uploads he can view the history in future.

We can also create, add, delete some users and also mind their activities. The user can add photos to it, can view it, but cannot have the right to delete or remove photos. It's also a secured search over the network.

V.CONCLUSION

How to effectively utilize the rich user metadata in the social sharing websites for personalized search is challenging as well as considerable. We propose a framework to exploit the users social activities for modified image search, such as observations and the involvement of interest groups. The query relevance and user preference are all together integrated into the final rank list. An experiment on a larger scale Flickr dataset shows that the proposed framework greatly out performs the baseline.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 2, Febraury 2014

VI.FUTURE ENHANCEMENT

Under the concept of heterogeneous image rich information network, many future works are there. It will be interesting to see how such kind of network structure may benefit various image mining and computer idea and responsibilities, such as categorization, and segmentation of images, Tag explanation, and collaborative filtering. As for the proposed algorithm a method named Content-Based Relevance Response (CBRR) is used to achieve the high retrieval quality by using the discovered patterns. 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 web scale search engine, a distributive computing extension will be considered. Considering dynamic environment is very important. One of the solutions 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 networks.
- Third, when a new image comes, update the sub networks.

REFERENCES

[1] 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.

[2] 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.

[3] 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.

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

[5] R.C. Veltkamp and M. Tanase, "Content-Based Image Retrieval Systems: A Survey," technical report, Dept. of Computing Science, Utrecht Univ., 2002.

[6] H. Tamura and N. Yokoya, "Image Database Systems: A Survey," Pattern Recognition, vol. 17, no. 1, pp. 29-43, 1984.

[7] W.I. Grosky, "Multimedia Information Systems," IEEE Multi- Media, vol. 1, no. 1, pp. 12-24, Spring, 1994.

[8] V.N. Gudivada and V.V. Raghavan, "Content-Based Image Retrieval Systems," Computer, vol. 28, no. 9, pp. 18-22, Sept. 1995.

[9] 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.

[10] 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.

[11] 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-combinatio n.pdf, 2008.

[12] S. Sclaroff, M. La Cascia, and S. Sethi, "Unifying Textual and Visual Cues for Content-Based Image Retrieval on the World Wide Web," Computing Vision Image Understanding, vol. 75, pp. 86- 98, July 1999.

[13] Z. Ye, X. Huang, Q. Hu, and H. Lin, "An Integrated Approach for Medical Image Retrieval through Combining Textual and Visual Features," Proc. 10th Int'l Conf. Cross-Language Evaluation Forum: Multimedia Experiments (CLEF '09), pp. 195-202, 2010.

[14] G. Jeh and J. Widom, "SimRank: A Measure of Structural-Context Similarity," Proc. Eighth Int'l Conf. Knowledge Discovery and Data Mining (KDD '02), 2002.

[15] L. Page, S. Brin, R. Motwani, and T. Winograd, "The Pagerank Citation Ranking: Bringing Order to the Web," technical report, Stanford InfoLab, 1999.

[16] D. Lizorkin, P. Velikhov, M. Grinev, and D. Turdakov, "Accuracy Estimate and Optimization Techniques for Simrank Computation," VLDB Endowment, vol. 1, no. 1, pp. 422-433, 2008.

[17] D. Fogaras and B. Racz, "Scaling Link-Based Similarity Search," Proc. 14th Int'l Conf. World Wide Web, pp. 641-650, 2005.

[18] J. Wang, H.-J.Zeng, Z. Chen, H. Lu, L. Tao, and W.-Y. Ma, "ReCoM: Reinforcement Clustering of Multi-Type Interrelated Data Objects," Proc. 26th ACM Ann. SIGIR Conf. Research and Development in Information Retrieval, pp. 274-281, 2003.

[19] X. Yin, J.H., and P.S. Yu, "LinkClus: Efficient Clustering via Heterogeneous Semantic Links," Proc. 32nd Int'l Conf. Very Large Data Bases, pp. 427-438, 2006.

[20] T. Deselaers, D. Keysers, and H. Ney, "Features for Image Retrieval: An Experimental Comparison," Information Retrieval, vol. 11, no. 2, pp. 77-107, 2008.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 2, Febraury 2014

BIOGRAPHY

Ms.R.Keerthika Completed Bachelor's Degree at SSM College of Engineering, India in the year 2007 and Master's Degree at Anna University of Technology Coimbatore, India in the year 2011.Registered Ph.D in 2012 at Anna University Chennai, India. She is having 7 years of experience in teaching in engineering, Currently she is working as Assistant Professor, Department of Information Technology at Karpagam College of Engineering. Her area of interest includes Data Mining, Image Processing, Wireless sensor Networks. She is a member in professional societies like ISTE, ACM, CSTA, IACSIT, ICGST, ISOC, IAENG and IRED.

Dr.C.Nalini Completed her Master's Degree at Bharathiyar University, India in the year 2000. Completed her Ph.D in 2011 at Anna University Chennai. She is having 20 years of experience in teaching in engineering, currently she is working as a Professor, Department of Information Technology at Kongu Engineering College. Her area of interest includes Data Mining, Network security, Database Management System. She is a member in professional societies like ISTE, CSI.