International Journal on Emerging Technologies (Special Issue NCETST-2017) **8**(1): 470-474(2017)
(Published by Research Trend, Website: www.researchtrend.net*)*

ISSN No. (Print) : 0975-8364
ISSN No. (Online) : 2249-3255

Optimization of mining technique of an Image by Image Content

Prashant Rajput and Prof. K S Vaisla
Associate Professor, Department of Computer Science,
Amrapali Group of Institutes, Shiksha Nagar, Lamachaur, Haldwani, Uttarakhand
Department of Computer Science Engineering, BTKIT, Dwarahat

ABSTRACT: In this paper we present a hypothesis for image retrieval system by image contents that take an image as the input query and retrieves images based on image features i.e. Color, Shape, and Texture. Image Retrieval is an approach for retrieving semantically-relevant images from an image database based on Automatically-derived image features. The unique aspect of the system is the utilization of function which gives single value for all the feature matrices of an image. The proposed procedure consists of two stages. First, here we are going to extract the image features and storing into the image server and then the user retrieval of that image.

Keywords: Extraction and Matrix Conversion function (EMC Function), Query Server, Image Server, Matrices, Vector.

I. INTRODUCTION

Data mining is an AI powered tool that can discover useful information within a database that can then be used to improve actions. To appreciate why businesses are so excited about data mining, you need only imagine that a major department store chain is looking for ways to boost sales. They have a large database containing information about customers and the nature of their purchases (with particulars such as identity of items, price, date, and time of sale). Suppose a data mining utility unearthed a pattern in the data which indicated that customers who shopped on Saturday afternoons and who made their initial

Purchase of the day in the shoe department tended to make, on average, 4 additional purchases from other departments and that the average member of this group spent more per visit than the typical shopper. Can you now envision the sort advertising campaign that the department store chain might want to embark upon?

A golden vein- Computing Analysis of customer information, better known as data mining, is finally delivering on its promises – and expanding into some promising new areas. The Economist Technology Quarterly (June 10, 2004). "In the old days, knowing your customers was part and parcel of running a business, a natural consequence of living and working in a community. But for today's big firms, it is much more difficult a big retailer such as Wal-Mart has no chance of knowing every single one of its customers .So the idea of gathering huge amounts of information and analysing it to pick out trends indicative of customers wants and needs data mining has long been trumpeted as a way to return to the intimacy of a small town general store. But for many years, data mining claims were greatly exaggerated.

In recent years however improvements in both hardware and software and rise of the World Wide Web, have enabled data mining to start delivering on its promises" from the Data Mining to Knowledge Discovery in Databases by Usama Fayyad, Gregory Piatetsky Shapiro and Pandhraic Smyth. AI magazine 17(3): Fall 1966, 37-54."Data mining and knowledge discovery in databases have been attracting a significant amount of research, industry and media attention of late .What is all the excitement about? This article provides an overview of this emerging field, clarifying how data mining and knowledge discovery in databases are related both to each other and to related fields, such as machine learning, statistics and databases. The article mentions particular real world applications, specific data mining techniques, challenges involved in real world applications of knowledge discovery and current and future research directions in the field.

By John Boyd, IBM thinks Research (1990). "Ordinary data mining simply looks for keywords, but the text mining system dubbed TAKMI (an abbreviation for text analysis and knowledge mining but also a Japanese word meaning 'skilled craftsman') spots grammatical relationships as well. Knowing which word is the subject, which is the web and which the object, TAKMI can categorize calls according to whether they are, say, complaints or questions and according to the product that is causing difficulty." Knowledge discovery and

data mining research at IBM. "The challenge of extracting knowledge from data draws upon research in statistics databases, pattern recognition, machine learning, data visualization, optimization, and high-performance computing, to deliver advanced business intelligence and web discovery solutions."

A. Data mining and complex objects

For many standard applications, like market analysis, constructing a usable KDD process is a rather well determined task. However, the data to be processed in real world applications is getting more and more complex and is yielding more potential knowledge. With advancing processors, memory and disc space, the detail level of objects is increasing as well as their plain numbers. For example, companies acquire more detailed information about their costumers, sky telescopes offer pictures with higher resolutions and html documents use structural tags, embedded multimedia content and hyperlinks which make them more complicated than ordinary text documents.

All these additional information yields new challenges to KDD. Though it is basically desirable to have more informa -are used in data mining gets more difficult. Additionally, many complex objects provide structural information as well as plain features. **For example**, a gene sequence is characterized by the order of nucleotides instead of their plain appearance in the gene.

To analyse complex objects, the most established way is to map any complex object to a feature vector. The idea is to span a vector space in which each relevant object characteristic or feature provides a dimension. Thus, an object is represented by the vector of its feature values. Since this is the most common feature representation, there is a wide variety of data mining algorithms that can process vectors as input representation. Though this method offers good results in many application areas, the data transformation becomes more and more difficult with increasing object complexity. Since data transformation usually is not informed about the purpose of the KDD task, it is difficult to decide which characteristic of an object should be preserved and which can be neglected. Furthermore, structural information is very difficult to express using a single feature vector. For example, it is not possible to model an arbitrary sized set within a feature vector without losing information. Thus, transforming complex objects into a feature vector and employing vector based data mining often spends large efforts for data transformation and provides suboptimal results.

For several applications, it is more beneficial to employ specialized data mining algorithms that can process more complex input representations than plain feature vectors. Employing structured object representations like graphs, sequences or relational data, often provides a more natural view on real world complex objects. The type of data representation discussed in this work is called compound object representation and is also capable to model structural information.

B. Complex and Compound Data Objects

Compound data objects are built of concatenations and sets of other compound data objects. Basic compound objects can consist of any object representation that can be processed by a data mining algorithm.

II. REVIEW OF LITERATURE

Data mining - Traditionally, algorithms for data analysis assume that the input data contains relatively few records.Currents databases, Current databases, however, are much too large to be held in main memory. Retrieving data from disk ismarkedly slower than accessing data in RAM. Thus to beefficient, the data-mining techniques applied to very largedatabases must be highly scalable. An algorithm is said to be*scalable* if-given a fixed amount of main memory – its runtimeincreases linearly is said to be scalar if – given a fixed amount ofmain memory –its runtimes increases linearly with the numberof records in the input database.

"Data mining, the ability to find unexpected patterns in accumulated data, was born during a lunch break. At a customer conference in the early 1990s, an executive at British departmentstore chain Marks & Spencer was explaining his database woesto Rakesh Agrawal, an information retrieval specialist at IBM. The store was collecting all sorts of data but didn't know what todo with it. So Agrawal and his team began devising algorithms for asking open-ended queries, eventually authoring a 1993paper that would become required reading in data-miningscience. The report has been cited in more than 650 otherstudies, making it one of the most widely cited papers of itskind. Agrawal, the data-mining pioneer, is today working on asystem that will scramble customer data in a way that will allow companies to study buying trends or other patterns while preserving strict privacy.'

Retrieving Complex Object (Like Images by Image Content)

- The application scenario is that the user inputs a rough sketchdepicting the prominent edges or contours of objects and wishesto retrieve database images that have similar shapes. We canonly expect to get a rough query sketch from the users, which islikely a distorted version of the intended database image, henceit is imperative

that tolerance be provided towards sketchdistortion ; by yin Chan and S.Y. Kung, Princeton University.

Because automated image retrieval is only meaningful in itsservice to people, performances characterization must begrounded in human evaluation of retrieval result, both for queryby images example and query by image example and query bytext. The data is independent of any particular image retrievalalgorithm and can be used to evaluate and compare many suchalgorithms without further data collection; by Nikhil VShirahatti, Kobus Barnard, University of Arizona.

III. RELEVANCE OF PROPOSED STUDY

In tradition 'search –and - retrieval' systems, users throughspecific queries to collections of text and get back more or lessuseful answers to those queries in text from again ,today inWWW era, we are dealing with images and video data at a largeextent. So the goal of data-mining should include large andcomplex object s like images or videos as well as text mining toproduce new knowledge by exposing unanticipated similaritiesor differences ,clustering or dispersal, co-occurrence and trendson large object also. With its roots in statistics, artificialintelligence and machine learning, data-mining has been aroundsince the 1990s.
The study would be very relevant towards identifyingimages on the basic of image content. Because now adays theimage data base increasing tremendously day by day.

IV. OBJECTIVES OF THE PROPOSED STUDY

This study will discuss the concepts related to the datamining technique used to solve the query for large and complexobjects. Data miners will often try different algorithms andsettings, and inspect the resulting models and test results toselect the best algorithm and settings. This study will provide ahigh-level overview of the algorithms supported by manystandards like classification problems: *decision tree, naïve bayes(NB), support vector machine (SVM), and feed forward neuralnetworks.*
In traditional 'search-and-retrieval' projects, scholars bringspecific queries to collections of text and get back more or lessuseful answers to those queries, 'By contrast, the goal of datamining,including text-mining, is to produce new knowledge byexposing unanticipated similarities or differences, clustering ordispersal, co-occurrence and trends.' With its roots in statistics, artificial intelligence and machine learning, data-mining hasbeen around since 1990s. With data-mining tools, you first selecta body of material that you think is important in some way, nextselect features of those materials that you similarly think are important, and then 'map the occurrence of those features in the selected materials to see whether patterns emerge. If patterns doemerge, you analyse them and from that analysis emerges if you are lucky-new insights into the materials.'

V. METHODOLOGY FOR THE PROPOSED STUDY

This hypothesis will discuss the concepts related to the datamining technique used to solve the query for large and complexobjects like images and videos. Data miners will often trydifferent algorithms and settings, and inspect the resultingmodels and test results to select the best algorithms and settings.
This study will provide a high-level overview of the algorithmssupported by many standards like classification problems,*decision tree, naïve bayes (NB), support vector machine (SVM),and feed forward neural networks.*

A. Proposed Architecture

Proposed function:
We suggested a mathematical function named as Extraction andmatrix conversion function which takes the resized image asinput and extract all the features (Shape, colour and texture) interm of their respective matrices and convert these matrices ofthe image in to the single and unique value correspondingly.
The steps involved in the function are as follows:
Step 1: Extract Features of resized image (Colour, Texture, andShape).

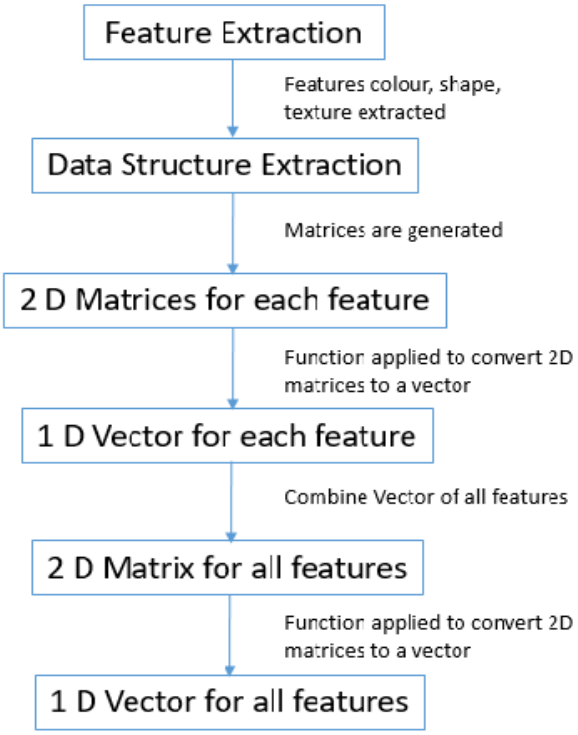

Fig.1. Feature extraction and matrix conversion.

Step 2: Extraction of Data Structure and 2D Matrices or eachfeature are generated.
Step 3: A Function is applied to convert the 2D matrices to asingle value for all different matrices.
Step 4: All single values of a particular feature (like shape,texture and colour) are combined to form 1D vector.
Step 5: All 1D vectors (of shape, colour and texture) arecombined to form a 2D matrix.

Properties of the Function

The function will give the single unique value for the combination of the different feature matrices.
The proposed architecture consists of two subsystems-
1. Image's Feature Extraction and storage, and
2. Retrieval of image.

The mathematical function is applied for the extraction process,and has the constraint that all the images stored of the same size(e.g.3x4cm). The process is shown in figure 1 and 2.

Steps for Image's Feature Extraction and Storage

Step 1: Original image is resized to size 3x4 cm.
Step2: Extraction and matrix conversion function is applied.
Step 3: The function generates the single and unique valuevector.
Step 4: This single and unique value is stored in the query serverwith reference to the image.

Retrieval of Image by the User

Step1: User queries for the Image.

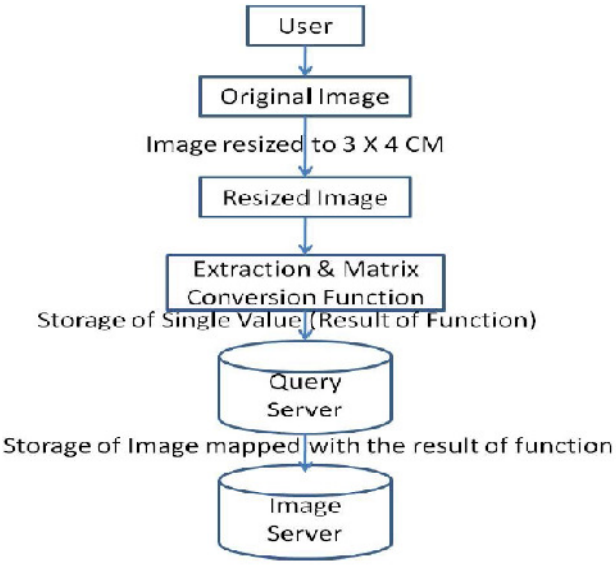

Fig. 2. Extraction of the Image Features and Storage of the Image with mapped value generated by function.

Step2: Image is resized to the fixed size (e.g. 3x4cm).
Step3: Extraction and matrix conversion function is applied toimage and get a single unique value for the image
Step4: This single value is compared with the image values stored in the query server and the corresponding image is displayed.

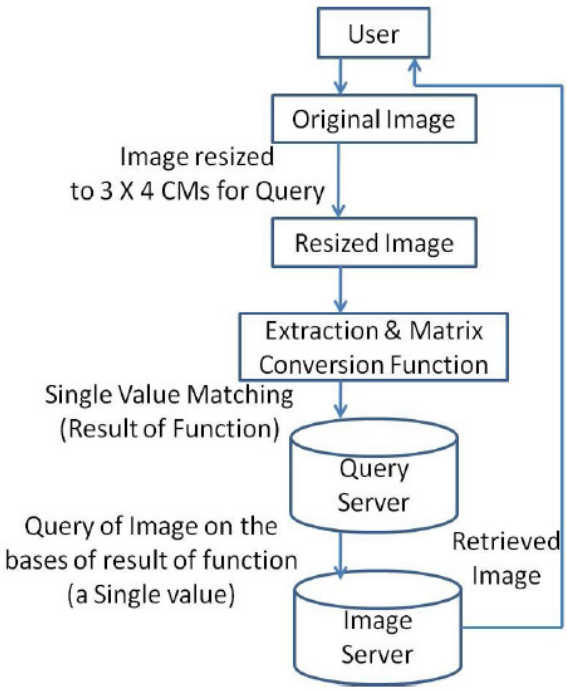

Fig. 3. Retrieval of the Image from Image Server with help of generated function value.

VI. CONCLUSIONS

The proposed system is a hypothesis; it provides a single unique value for the combination of the different features matrices. It makes the searching for the image retrieval faster and more efficient because we compared the single value for the different features (colour, shape, and texture) matrices. Due to this mathematical function the proposed system increases the searching efficiency up to 60-70%.

VII. SCOPE AND LIMITATIONS

As this is hypothesis and most of the part of the proposed system shall be carried out theoretically, practical of research work shallnot be covered in this work.

REFERENCES

[1]. Performance mining of large-scale data-intensive applications, Carothers, C. Szymanski, B.K.; M. Paralleland Distributed Processing Symposium, Proceedings International, IPDPS 2002, Abstracts and CD-ROMvolume, Issue, 2002 Page(s):177-178.

[2]. Advanced Scout: Data Mining Knowledge Discovery inNBA Data, a Brief Application Description. By Inderpal Bhandari, et al. Data Mining and knowledge Discovery1,121-125(1997).

[3]. Mining very Large Databases, Venkatesh Ganti,Johannes Gehrke, Raghu Ramakrishnan; university of Wisconsin-Madison, IEEE Computer , PP 38-45, 0018-9162/99-@ August 1999.

[4]. "Distance Measures for point Sets and TheirComputation". T. Eiter and H.Mannila. Acta Information,34(2):103-133,1997.

[5]. Google Press Center. Google Archives search Milestonewith Immediate Access To More Than 6 Billion Items.

[6]. Knowledge Discovery and Data Mining: Towards a unifying framework. U.M. Fayyad, G. Piatetsky-Shapiro,and P. Smyth. In Knowledge Discovery and Data Mining, Pages 82-88, 1996.

[7]. "A polynomial time computation metric between pointssets". J. Romon and M. Bruynooghe. Acta Information,37:765-780,2001.

[8]. HIERARCHICAL ALGORITHM FOR IMAGERETRIEVAL BY SKETCH, Yin Chan and S.Y. Kung , Princeton University pp-564-569, 0-7803-3780-8/97/$10.00 01 997 IEEE.

[9]. Evaluating Image Retrieval , Nikhil V Shirahatti, KobusBarnard , Proceedings of the 2005 IEEE computer Society Conference on computer vision and Pattern Recognition(CVPR'05) 1063-6919/05@ 2005 IEEE.

[10]. The Rebirth of Artificial Intelligence. Lisa DiCarlo.Forbes (May 16, 2000). "Oracle is promoting itsIntelligence Web House tools.