2018-19

Improved Performance in CBIR using Machine Learning Approach

B. Ramesh Naik
Asst Prof, Dept of CSE
GITAM School of Technology
GITAM (Deemed to be University)
Bengaluru, India
rameshnaik3@gmail.com

T. Venugopal
Professor, Dept of CSE
College of Engg, JNTUH
Jagityal, Karimnagar, India
t_vgopal@rediffmail.com

Abstract—We proposed a new approach which improves the high level image semantics based on machine learning approach. The contemporary approaches for image retrieval and object recognition includes Fourier transforms, Wavelets, SIFT and HoG. Though these descriptors helpful in a wide range of applications, they exploits zero order statistics and this lacks high descriptiveness of image features. These descriptors usually take benefit of primitive visual features such as shape, color, texture and spatial locations to describe images. These features do not adequate to describe high level semantics of the images. This leads to gap in semantic content caused to unacceptable performance in image retrieval system. A novel method has been proposed referred to as discriminative learning which is derived from machine learning approach that efficiently discriminates image features. The analysis and results of proposed approach were validated thoroughly on WANG and Caltech-101 Databases. The results provide evidence that this approach is very competitive in Image retrieval system.

Keywords—CBIR, Discriminative Learning, Region Weight Learning, SIFT.

I. INTRODUCTION

Content Based Image Retrieval (CBIR) is essentially image retrieval system where specified number of similar images are retrieved from large image database based on primitive visual features and derived semantics to a given query image. CBIR gained significant importance in last 15 year in which colour, shape, texture and spacial locations are used as primitive visual features to describe images. A great research has been attracted to characterize primitive properties of the image in past 10 years. The task of primitive feature extraction plays a elementary role for the efficiency of middle-level or high-level descriptors in the vision tasks. The tricky task in CBIR is to develop descriptors which are high distinctiveness i.e features are invariant to scale, translation and rotation for general image retrieval tasks. Diverse image representation techniques have been invented for extracting primitive features in the query and target images localized by region detectors, shape detectors or color gradients detectors etc. [21], [25], [29]. A CBIR system is described in [2], which utilizes Gabor texture features as descriptors for image retrieval. Texture features are extracted by the statistical quantities mean, variance of Gabor filtered image. The rotation normalization in their approach is achieved by concept of circular shift as a result every image encompasses

identical dominant direction. A Texture based transformation approach such as second order-moment matrix and multi-orientation filter banks are used in [1], [23] to represent image regions which are coherent in image retrieval system. They also used Gaussian window and Expectation-Maximization method to accomplish better image retrieval accuracy. A novel framework is proposed in [3] which combine texture, color and shape information to achieve higher retrieval accuracy in CBIR. In their approach each image is divided into set of segments and low level features color and texture of these segments are extracted from resulting of Gabor filtered image. Machine learning model is developed in [5] which discover some structures which are resulting from Gaussian process (GP) task. These problems are progressively more frequent and important, but have until then presented severe practical difficulties. Finally they introduce a new method for Gaussian Process approximation, which yield boundary forecast robust to hyperparameter specification. Venugopal et.al [4] presented a novel method for CBIR using Adapted Fourier Descriptors to describe shape features which are invariant to scale, rotation, translation of operated images. A CBIR system proposed in [9] using Wavelet Transformation and Deep Autoencoder method. Wavelets transforms is potential technique for texture based image retrieval and describes texture features of the given image efficiently. Montazer et.al [10] proposed combined approach using SIFT and wavelet transform to efficiently depict image features. Peihua et.al [6], [7] used multivariate Gaussian distribution method for image classification and also described the process of embedding Gaussian space into linear space by defining a operation in terms of multiplication on the manifold. A Generic Fourier descriptor approach presented in [8] to uncover the limitations of existing shape description techniques in shape based image retrieval. In their approach a 2-D Fourier transform is applied on shapes to determine features.

The prime factor in Content-Based Image Retrieval is building and organizing image databases that influence efficiency of the retrieval system. CBIR [4], [12] is a essential applications of computer vision task in which visual and derived features are used as key components for matching the images in large databases. The state of the art approaches invented for image representation, retrieval and object recognition which includes Fourier transforms, Wavelets, SIFT and HoG. These descriptors are applied in several

2018-19

RATIO DELAY STUDY OF PUBLIC TRANSPORT BUSES

[1]Mr.M.Krishna,

[1]Research Scholar at JNTUH & Asst.Professor, Matrusri Engineering College, Saidabad, Hyderabad-

[2]Dr. N.V.S. Raju

[2]Professor, College of Engineering, JNTUHJ, Nachupally, Jagithyal, Telangana, India.

ABSTRACT : The top management of any organization, either private or government always want to know the utilization of their resources i.e. manpower or machines or financial resources so that they can initiate necessary actions for maximum utilization and hence higher productivity for their organization. Work sampling or Ratio delay study is one of the techniques to identify the real situation of whether the workers or manpower are working or not working at a given point of time.

The proposed work is carried out at Telangana State Road Transport Corporation (TSRTC), which is into Public bus Transportation in Telangana and other states. The objective of this paper is to assess the percentage of working of buses for a given period and the loss incurred due to non operation. The data is collected, from one of the Depots of the Corporation. This analysis will help in understanding the failure rate of buses and the management may think of refining the maintenance schedules to improve the service level to customers.

Key Words: **Activity sampling or Ratio delay study, work measurement, productivity, probability, random sampling**

I. INTRODUCTION

Activity sampling was developed by L.H.C. Tippet in Britain in 1934 for the British Cotton Industry Research Board. R.L.Morrow used this technique in America during 1945 and named it as Ratio Delay Study. In 1952, C.L.Brisley renamed this technique as Work sampling. Presently, it is the most commonly used technique of Work Measurement.

Activity sampling as defined by B.S.3138: 1969, is a technique in which large number of observations are made over a period of time of one or group of machines, processes or workers. Each observation records what is happening at that point of time or instant and the percent of observations recorded for a particular activity or delay is a measure of the percentage of time during which that activity or delay occurs.

It is a technique for analyzing activity by taking a large number of observations at random times and to investigate the proportions of total time devoted to the various activities that constitute a job or work situation. The results of work sampling are effective for determining machine and personnel utilization, allowances applicable to the job and production standards. It provides the same information faster and considerably at less cost than time study.

The ratio of observations of a given activity to the total observations approximates the percentage of time that the process is in that state of activity. The accuracy of the data determined by work sampling depends on the number of observations and the period over which the random observations are taken.

Advantages of Work Sampling
(i) It does not require continuous observation by an analyst over a long period of time,
(ii) Clerical time is diminished,
(iii) The total work-hours expended by the analyst are usually much fewer,
(iv) The operator is not subjected to long period stopwatch observations.
(v) Crew operations can be readily studied by a single analyst.

Uses of Work Sampling:
(i) Activity and delay sampling (to measure the activities and delays of workers or machines. Ex: Determine machine utilization)
(ii) Perfomance sampling (to measure working time and nonworking time of a person on a manual task and to establish a performance index or performance level for the person during his or her working time).
(iii)Work measurement (under certain circumstances, to measure a manual task, that is, to establish a time standard for an operation.

II. THEORY OF WORK SAMPLING (RATIO DELAY STUDY)
The theory of work sampling is based on the fundamental law of probability. At a given instant, an event can be either present or absent. Random sampling requires that there be no bias in the sampling process. It is important that the concept of randomness be understood and carefully followed in work sampling studies.

The normal distribution curve is typical of the kind of frequency distribution which is of importance in work sampling because it represents graphically the probabilty of the occurence of certain chance phenomena. At the outset it is necessary to decide what level of confidence is desired in the final work sampling results.

Image and Video Clustering Techniques in High Dimensional Data Set - A Review

Mrs. D. Manju	Dr. M. Seetha	Dr. P.Sammulal
Research Scholar,	HOD & Professor,	Associate Professor,
Dept of CSE, JNTUH	Dept.of CSE, GNITS,	Dept. of CSE, JNTUH
jadavmanju@yahoo.co.in	maddala2000@yahoo.com	sam@jntuh.ac.in

Abstract -- Clustering deals with grouping of unlabeled data, having wide applications in computer vision, data mining, data visualization, etc. Many computer vision problems require the efficient and effective organization of huge-dimensional data for information retrieval, indexing and searching of Image or Video Datasets. This paper attempts to present methods for Image and video clustering on High Dimensional Datasets based on runtime. The algorithms included in the study are Histogram clustering, Hierarchical clustering using random projections, Mean-shift formulation and Histogram of textons. The Histogram clustering represents the distribution of the values of an attribute. Hierarchical clustering using random projections describes the recursive nature of algorithm. Mean-shift formulation merges the samples during the spatiotemporal mean-shift procedure and groups the similar clusters after convergence. Histogram of textons is used to recognize textured patches that construct textons across a variety of textures. The above algorithms are compared and analyzed with respect to run time .The comparison results shows that 1D Random Projection's maximum runtime is even smaller than the minimum runtime of PROCLUS method. Hence, it was ascertained that Hierarchical clustering using random projections is extremely efficient for image clustering.

Index Terms -- Clustering, Clustering, Histogram, Key Frames, Shots, unsupervised.

I. Introduction

Clustering uses unsupervised learning methodology for grouping datasets on distance measurements, density measurements or statistical distributions. The growth of digitally stored images and video datasets has increased the need for tools to search and organize visual data automatically by their content. Browsing environments help users to find the images or videos they need by structuring images and videos. Therefore, for grouping semantically related images or videos, clustering techniques are required which are highly scalable, and produce balanced structures that are useful for content based browsing. These factors depend on the type of clustering model used. Analysis of an image can be manipulated in a number of ways. Image databases contain images with different characteristics, to cluster image databases extract features from an image and apply clustering techniques. Selecting a feature to an algorithm is an intricate task. The aim of Video analysis is to obtain a structured organization of the original video content and understanding its

Sentiment-Based Semantic Rule Learning for Improved Product Recommendations

Dandibhotla Teja Santosh and
Bulusu Vishnu Vardhan

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.72514

Abstract

Crucial data like product features and opinions that are obtained from consumer online reviews are annotated with the concepts of product review opinion ontology (PROO). The ontology with instance data serves as background knowledge to learn rule-based sentiments that are expressed on product features. These semantic rules are learned on both taxonomical and nontaxonomical relations available in PROO ontology. These rule-based sentiments provide important information of utilizing the relationship among the product features 'as-a-unit' to improve the sentiments of the parent features. These parent features are present at the higher level near the root of the ontology. The sentiments of the related product features are also improved. This approach improves the sentiments of the parent features and the related features that eventually improve the aggregated sentiment of the product. The result is either the change in the position of the product in the list of similar products recommended or appears in the recommended list. This helps the user to make correct purchase decisions.

Keywords: recommender system, product feature, feature sentiment, ontology, rule-based sentiments, purchased decision

1. Introduction

Traditional machine learning algorithms experience the data and learn the hypothesis. Tree and rule-based algorithms learn the hypothesis using the attribute-value pairs from the input data. Machine cannot go beyond the task of identifying features and opinions from the reviews as it never possess prior knowledge to understand the relationships among the attributes and context specific constraints that are available among the product features and opinions.

Semantic web ontology helps to overcome this problem. Ontology [1] encodes the relationships among the concepts of features and opinions with inequality constraints, semantic

Gender Prediction in Author Profiling Using ReliefF Feature Selection Algorithm

T. Raghunadha Reddy, B. Vishnu Vardhan, M. GopiChand
and K. Karunakar

Abstract Author Profiling is used to predict the demographic profiles like gender, age, location, native language, and educational background of the authors by analyzing their writing styles. The researchers in Author Profiling proposed various set of stylistic features such as character-based, word-based, content-specific, topic-specific, structural, syntactic, and readability features to differentiate the writing styles of the authors. Feature selection is an important step in the Author Profiling approaches to increase the accuracy of profiles of the authors. Feature selection finds the most relevant features for describing the dataset better than the original set of features. This is achieved by removing redundant and irrelevant features according to important criteria of features using feature selection algorithms. In this work, we experimented with a ReliefF feature selection algorithm to identify the important features in the feature set. The experimentation carried on reviews domain for predicting gender by using various combinations of stylistic features. The experimental results show that the set of features identified by the ReliefF feature selection algorithm obtained good accuracy for gender prediction than the original set of features.

T. Raghunadha Reddy (✉) · M. GopiChand
Department of IT, Vardhaman College of Engineering, Hyderabad, India
e-mail: raghu.sas@gmail.com

M. GopiChand
e-mail: gopi_merugu@yahoo.com

B. Vishnu Vardhan
Department of CSE, JNTUH Jagtiyal, Karimnagar, India
e-mail: mailvishnu@jntuh.ac.in

K. Karunakar
Department of CSE, Swarnandhra Institute of Engineering and Technology, Narsapur,
Andhra Pradesh, India
e-mail: karunakar.mtech@gmail.com

169

An Optimal Heuristic for Student Failure Detection and Diagnosis in the Sathvahana Educational Community Using WEKA

P. Vasanth Sena and Porika Sammulal

Abstract The study offered in this paper aims to explore students characteristics and to determine unsuccessful student groups in respective subjects based on their earlier education and the impact of other factors in multiple dimensions. Predictive data mining techniques such as as classification analysis is applied in the analysis process. Datasets used in the investigation were collected from all academic years in the Sathavahana educational community contains different professional disci plines through online. The method adopted is to know the number of students failing in each subject and analyze the reasons for failure using data mining tools like WEKA. This model works effectively with large datasets. It has been tested on WEKA with different algorithms.

Keywords Failure detection · Diagnosis · Educational · Institutions Classification

1 Introduction

Every academic year thousands of students are enter higher educational institutions. As an example, in our Sathavahana educational institutions at least two thousand students enrol every academic year under different disciplines. In the traditional methods, in order to maintain student's details and analysis of pass percentage, precautions to reduce the failure and detentions percentage is difficult, and then the data set exponentially increases.

Existing methods, such as classroom attendance registers for monitoring attendance period-wise, and some computer Excel-based formats that maintain details

P. Vasanth Sena (✉)
R&D Cell, JNTUH, Hyderabad, India
e-mail: vasanthmtechsit521@gmail.com

P. Sammulal
CSE Department, JNTUH-CEJ, Karimnagar, India

671

Association Rule Mining Using an Unsupervised Neural Network with an Optimized Genetic Algorithm

Peddi Kishor and Porika Sammulal

Abstract The best known and most widely utilized pattern finding algorithm in data mining applications is association rule mining (ARM). Extraction of frequent patterns is an indispensable step in ARM. Most studies in the literature have been implemented on the concept of support and confidence framework utilization. Here, we investigated an efficient and robust ARM scheme based on a self-organizing map (SOM) and an optimized genetic algorithm (OGA). A SOM is an unsupervised neural network that efficaciously produces spatially coordinated internal feature representations and detected abstractions in the input space and is the most efficient clustering technique that reveals conventional similarities in the input space by performing a topology maintaining mapping. Hence, a SOM is utilized to generate accurate clustered frequency patterns and an OGA is used to generate positive and negative association rules with multiple consequences by studying all possible patterns. Experimental analysis on various datasets has shown the robustness of our proposed ARM in comparison to traditional rule mining approaches by proving that a greater number of positive and negative association rules is generated by the proposed methodology resulting in a better performance when compared to conventional rule mining schemes.

Keywords Data mining · Association rule mining · Frequent patterns
Positive and negative association rules · Self-organizing maps (SOM)
Optimized genetic algorithm (OGA)

P. Kishor (✉)
R&D Cell, JNTUH, Hyderabad, India
e-mail: kishorpeddi25@gmail.com

P. Sammulal
Department of CSE, JNTUH College of Engineering, Jagtial, India

657

Impact of Term Weight Measures for Author Identification

M. Sreenivas, T. Raghunadha Reddy and B. Vishnu Vardhan

Abstract The rapidly growing data in the Web result in stolen, unidentified, and fraudulent data. Identification of such data is of a prime objective for forensic departments, researchers, and governments. In this context, authorship analysis is very useful to reveal the truth by analyzing the text. Authorship analysis is observing the properties of a text to predict authorship of a document. Stylometry is the root for authorship analysis, which is a linguistic research field that exploits the machine learning techniques as well as knowledge of statistics. Authorship Attribution is a type of authorship analysis technique, which is aimed at recognizing the author of an anonymous text within a closed set of authors or subjects. Most of the researchers in Authorship Attribution approaches proposed various set of stylistic features to differentiate the authors based on style of writing. It was observed from the literature the accuracy of author prediction was not satisfactory with stylistic features. In this paper, the experimentation carried out with various term weight measures identified in various text processing domains to predict the author of a new document. The results show that the term weight measures obtained good accuracies for author prediction when compared with most of the existing approaches.

Keywords Term weight measure · Author identification · Authorship attribution
BOW model

M. Sreenivas
Research Scholar of RUK, Sreenidhi Institute of Science and Technology, Hyderabad, India
e-mail: msreenivas@sreenidhi.edu.in

T. Raghunadha Reddy (✉)
Dept of IT, Vardhaman College of Engineering, Hyderabad, India
e-mail: raghu.sas@gmail.com

B. Vishnu Vardhan
Dept of CSE, JNTUH Jagtiyal, Karimnagar, India
e-mail: mailvishnu@jntuh.ac.in

81

A New Document Representation Approach for Gender Prediction Using Author Profiles

T. Raghunadha Reddy, M. Lakshminarayana, B. Vishnu Vardhan, K. Sai Prasad and E. Amarnath Reddy

Abstract Author Profiling is used to predict the demographic characteristics like age, gender, country, nativity language, and educational background of anonymous text by analyzing their style of writing. Several researchers proposed different types of features like lexical, character based, content based, syntactic, topic specific, structural features, and readability features to discriminate the style of writing of the authors for Author Profiling. The representation of a document with extracted features is one of the important tasks in Author Profiling. In Author Profiling approaches, most of the researchers used the bag-of-words model for document representation. This paper concentrates on the alternative document representation to increase the performance of Author Profiling system. In this work, a new document representation model is proposed and compared the proposed model with existing document representation models like BOW and SOA. The proposed model is evaluated on the reviews dataset for predicting the gender of the authors using various machine learning classifiers. The proposed approach results were promising than most of the existing approaches for Author Profiling.

T. Raghunadha Reddy (✉)
Department of IT, Vardhaman College of Engineering, Hyderabad, India
e-mail: raghu.sas@gmail.com

M. Lakshminarayana
Department of CSE, Swarnandhra College of Engineering and Technology, Narsapuram, AP, India
e-mail: lachi9866516918@gmail.com

B. Vishnu Vardhan
Department of CSE, JNTUH Jagtiyal, Karimnagar, India
e-mail: mailvishnu@jntuh.ac.in

K. Sai Prasad · E. Amarnath Reddy
Department of Computer Science and Engineering, MLR Institute of Technology,
Hyderabad, India
e-mail: saiprasad.kashi@gmail.com

E. Amarnath Reddy
e-mail: amar.enumula@gmail.com

39