Department of Collegiate Education

Government College for Women

Maddur - 571 428

Re-accredited by NAAC with 'B' Grade

Ph : 08232-236867, E-mail : gcwmaddur@gmail.com

One Day National Conference (Multidisciplinary)

On

"Emerging Trends in Digital Transformation - ETDT 2018"

30th July 2018

Volume 1

Organized by

IQAC, Department of Humanities, Science, Commerce & Management.

An Enriched framework on Sentiment Analysis of Movie Reviews Data Sets using Machine Learning Algorithm

K.Deepa* & Kirubakaran**

Abstract

In this research paper our target set is movie reviews. There are diverge range of mechanisms to express their data which may be either subjective, objective or a mixture of both. Besides the data collected from World Wide Web consists of lot of movie dataset. It is very much true that we are going to apply some pre-processing techniques and compare the accuracy using Machine Learning algorithm for various classifiers. With ever growing demand to mine the Big Data the open source software technologies.

Keywords: Sentiment Analysis, SVM, K-Nearest Neighbour, Random Forest, Naive Bayes Algorithm.

Introduction

Sentiment analysis refers to the use of natural language processing, text analysis and computational linguistics to extract and identify subjective information in source materials. It

aims to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document [1].

[2]In last decade there is a rise of social media such as blogs and social networks, which has fueled the interest in sentiment analysis. Online opinion has turned into a kind of virtual currency with the proliferation of reviews, ratings, recommendations and other forms of online expression, for businesses that are looking to market their products, identify new opportunities and manage their reputations. In order to automate the process of filtering out

the noise, understanding the conversations, identifying the relevant content and following appropriate actions, many are now looking to the field of sentiment analysis.

[3] The problem of most sentiment analysis algorithms is that they use simple terms to express sentiment about a Product or service. However, cultural factors, sentence negation, sarcasm, terseness, language ambiguity and differing contexts make it extremely difficult to turn a string of written text into a simple pro or con sentiment.

Data Pre-processing

The dataset provided in this competition is comprised of tab-separated files with phrases from the Rotten Tomatoes dataset [4]. The dataset has been divided to training and test set for the purpose of benchmarking, but the sentences have been shuffled from their original order. Each Sentence in the dataset has been parsed into many phrases by the Stanford parser. Each

Phrase has a phrase Id. Each sentence has a sentence Id. Phrases that are repeated (such as short/common words) are only included once in the data.

The train.tsv contains the phrases and their associated sentiment labels. The test.tsv contains Just phrases. Sentiment label to each phrase in test file should be assigned.

* Research Scholar, Bharthidasan University, Tiruchirappalli, India.deepamohan13@gmail.com,Mobile no: 9842129551

** Head, Department of Computer Science, Karunya University, Coimbatore, India.ekirubakaran@gmail.com Mobile no: 9442631150.

A Hybrid Method Using Rsa with Fast Exponentiation Algorithm

R. Kalaivani*

ABSTRACT— RSA encryption is one of the public-key methods that have been popular in last decade. Considering increment of security requirements, size of the keys has been larger. With key length growing, delay of exponentiation computation has changed into major problem in selecting longer keys. The necessity for information security becomes more widespread in these days, especially for hardware-based implementations such as smart cards chips for wireless applications and cryptographic accelerators. The binary or in other words square-and-multiply method is the classical exponentiation technique that is used in RSA. In this paper a new algorithm of exponentiation in RSA is presented that works in parallel, needs fewer multiplications and so has less delay. Therefore this technique is more useful in larger key computations. Fast modular exponentiation algorithms are often considered of practical significance in public-key cryptosystems.

KEYWORDS— Cryptography, Decryption, Encryption, Network Security.

I. INTRODUCTION

The RSA cryptosystem is one of the most widely used technologies for achieving information security. The main task of the encryption and decryption engine of RSA cryptosystem is to compute $M^E \bmod N$. Because the bit-length of the numbers M, E, and N would be about 512 to 1024 bits now, the computations for RSA cryptosystem are time-consuming. In this paper, an efficient technique for parallel computation of the modular exponentiation is proposed and our algorithm can reduce the time.

II. TYPES OF CRYPTOGRAPHY

Cryptography is one of the most popular techniques that can be used to provide sufficient security to the sensitive data. Cryptography is a method of storing and transmitting data in a particular form so that only those for whom it is intended can read and process it. There are several ways of classifying cryptographic algorithms. Secret Key Cryptography (SKC) uses a single key for both encryption and decryption .Public Key Cryptography (PKC) uses one key for encryption and another for decryption.

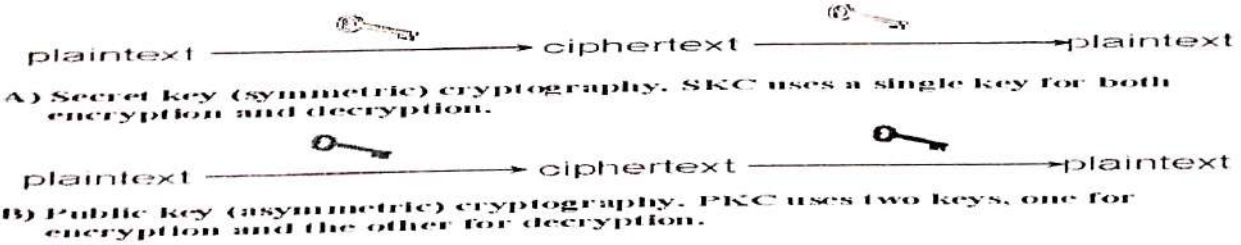


Fig 1: Types of cryptography

III. SYMMETRIC KEY CRYPTOSYSTEM

Symmetric encryption also referred to as conventional encryption or single key encryption was the only type of encryption.

Cipher text = encrypt (plaintext, key) **Plaintext = decrypt (cipher text, key)**

Fig 2 shows the symmetric encryption scheme with five ingredients such as **Plaintext** is the original intelligible message or data that is fed to the algorithm as input.The **Encryption Algorithm** performs various substitutions and permutations on the plaintext. The **Secret Key** is also input to the

* Assistant Professor, Department of CS, Nehru Memorial College, Puthanampatti, Trichy (DT). kalaivinayak@gmail.com

A Survey on Mining Frequent Itemsets

P.Sumathi* & S. Murugan**

Abstract - Data mining is a process of extracting valuable and hidden information from large databases. Association rule mining is one of the popular data mining techniques. In the field of association rules mining, Apriori is most popular algorithm which is used for finding frequent itemsets from large databases. But, the application of Apriori algorithm needs to scan the database many times which leads to a great overhead. In last decades, lots of research activities has been done to improve the efficiency of Apriori algorithm and this paper presents a comprehensive survey of the latest techniques and developments in finding frequent itemset mining and association rule mining.

Keywords - Data Mining, Frequent Itemsets, Association Rule Mining and Apriori Algorithm

Introduction

Association rule mining is one of the popular data mining techniques which have wide range of applications in marketing and retail communities as well as other more diverse fields [6]. Discovery of association rules involves two major steps. 1) Finding frequent itemsets and 2) generating reliable and strong association rules from the frequent itemsets. Step 1 of association rule mining is a challenging task [4] and plays a vital role in mining associations and correlations [5]. Association rules are generated for detecting unknown relationships, producing results which can perform basis for decision making and prediction. Researchers have proposed various improved algorithms for generating frequent itemsets. These algorithms differ in their ways by how they traverse and handle the database. This paper presents an overview of the latest innovations and improvements in finding frequent itemsets proposed by various researchers.

The rest of the paper is organization as follows. In Section 2, describes the Apriori Algorithm and its limitations. In Section 3 elaborates the works related to Apriori and its improvement and the conclusion is given in Section 4.

Apriori Algorithm

Apriori Algorithm is one of the most important algorithms which are used to extract frequent itemsets from large database and obtain the association rule for discovering the knowledge [1] and proposed by R. Agrawal and R. Srikant in 1994 [5]. Apriori employs an iterative approach known as a level-wise search, where k-itemsets are used to explore (k+1)-itemsets. First, the set of frequent 1-itemsets is found by scanning the database to accumulate the count for each item, and collecting those items that satisfy minimum support. The resulting set is denoted by L_1. Next, L_1 is used to find L_2, the set of frequent 2-itemsets, which is used to find L_3, and so on, until no more frequent k-itemsets can be found. The finding of each L_k requires one full scan of the database [15].

To improve the efficiency of the level-wise generation of frequent itemsets, an important property called the Apriori property i.e., all nonempty subsets of a frequent itemset must also be frequent is used to reduce the search space. To generate L_k from L_{k-1}, a two step process is followed. 1) Join step 2)Prune step. Join step generates C_k a set of candidate k-itemsets by joining L_{k-1} with itself and C_k is pruned to generate L_k with "minimum support" as the pruning parameter. To select the best rule among all the available rules, association rule mining depends on two important constraints. They are minimum threshold on support and confidence called minimum support and minimum confidence and is set by domain experts [12]. The support of an association rule is defined as the percentage of the fraction of records that contain to the total number of records in the database.

* Research Scholar, Department of Computer Science, Nehru Memorial College (Autonomous), Puthanampatti, Tiruchirappalli -Dt, TamilNadu, India, 9566385539,sumiparasu@gmail.com

** Associate Professor, Department of Computer Science, Nehru Memorial College (Autonomous), Puthanampatt, Tiruchirappalli -Dt, TamilNadu, India,9486631199,smurugan.nmc@hotmail.com

Enhanced Byte Rotation Encryption Algorithm Using Cipher Block Chaining Technique

Velmurugan P*

Abstract

In this digital world protecting the Sensitive data is a required thing. There is a demand for powerful cryptography techniques such as AES, DES, RSA, Blowfish, etc are used to encrypt the data. The quality of the encryption method is measured based on maintaining security and accuracy of the original message. Most of the algorithms achieve data security at greater level, but increasing security can leads to increase time and complexity of the algorithms. This may cause decreasing speed and efficiency of the encryption system. Enhanced Byte Rotation Algorithm (EBREA) using Cipher Block Chaining technique is the new method that ensures speed, security and accuracy of the encryption scheme. Byte Rotation Encryption Algorithm (BREA) works based on Symmetric a key cryptography and Block Cipher algorithm which divides the plain text into n number of plain text blocks using symmetric key technique but it has security issues. To overcome these flaws BREA is enhanced by combining it with Cipher Block Chaining (CBC) algorithm.

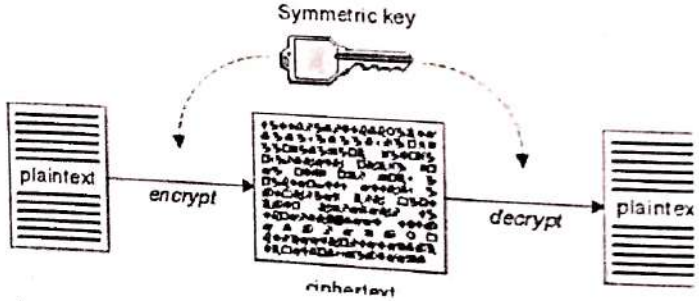
Key Words: Data, Security, Encryption, Decryption, Security, Cipher Block Chaining

I. INTRODUCTION

Cryptography is an art of exchanging the information in the secret manner. It has two main processes they are Encryption and Decryption. Encryption is done by the sender. During the encryption process, the original message (is called as Plain text) is converted into a cipher text by using mathematical operations and some additional information called as key. The generated cipher text has no meaning and no readable form. Then the Cipher text is transferred to the receiver through the network. The receiver receives the cipher text. The receiver decrypts the received cipher text by using any mathematical operations and additional information. This process is called Decryption. At the end of decryption the cipher text is to be converted into original plain text.

II. SYMMETRIC KEY AND ASYMMETRIC KEY CRYPTOGRAPHY

There are two types of crypto systems used to protect the data, commonly known as symmetric key cryptography and asymmetric key cryptography.



Symmetric Key Cryptography

In this cryptography both encryption and decryption process may take same key. This method is also called as private key cryptography.

Asymmetric Key Cryptography

In this cryptography sender and receiver uses different keys to encrypt and decrypt the data. The

* Assistant Professor, Dept of Computer Science, Nehru Memorial College, Puthanampatti. Mobile: 8110982923, Mail: vallivelprof@gmail.com

Review: Vertical Data Formats Of Frequent Pattern Mining Algorithms

V.Priya*, Dr. S.Murugan**

ABSTRACT: In the current digital world, there is a growth of data for everyday. The researchers are confused by the huge arrival of data. To manage all these massive data is an open challenge for all the researchers in frequent pattern mining. This paper gives a report of the concise narrate of earlier works of vertical formats of frequent patterns with a thorough description of the core algorithms, highlighting the major contributions of the different algorithms put forth by different authors. A broad review of the most significant algorithms of literature has been explained and compared with them in the vertical data Formats.

KEYWORDS: Frequent Pattern Mining, Association Rule Mining, Vertical data formats.

I. INTRODUCTION

Data Mining is the process of discovering knowledge. It is the process of extracting information from available raw data. The data are stored in databases. There are various kinds of data that can be used in data mining which includes transactional data, statistical data etc. Data mining includes various techniques for each purpose. Techniques include Association rule mining, classification, prediction, regression etc. Association rule mining techniques are widely used in discovering hidden correlations and relationship between the set of items in a transaction. It includes every transaction in the database during the discovery process. It also reveals the set of strategies that can be followed or neglected in the field for respective development .Many analysis data mining as synonymous to Knowledge Discovery from Data (KDD), while others consider data mining as an essential stage in the KDD process.

1.1 Data formats of fim algorithms

Frequent item set mining (FIM) is a useful tool for discovering frequently co-concurrent items. Since its inception, a number of significant FIM algorithms have been developed to speed up mining performance. Association rule mining is one of the most important data mining problems. The conviction of association rule mining is the discovery of association relationship among a set of items. There are two different data formats: horizontal data format and vertical data format. The data format of Eclat is vertical data format. Both data formats are the representations of the database in memory. Horizontal data format, a "TID-itemset" format, can also be expressed as "TID: itemset", in which "TID" is the unique identifier of a transaction in a database; "itemset" is the set of items included in the transaction. Vertical data format, an "Item-TID_set" format, can also be expressed as "Item: TID_set", in which "Item" is an item in database, "TID_set" is the set of transactions including the items. Table 1 shows the horizontal data format representation of a database and Table 2 shows the corresponding vertical data format representation. Two data formats have different efficiency in calculating the support degree of the itemset. When adopting the horizontal data format, it needs to scan the whole database to get the support degree of the itemset. But if adopting the vertical data format, it only needs to calculate the intersection of all TID_sets of items to get the support degree. For example, when we want to get the support of an itemset $X=\{I1,I2\}$, for the data in Table 1, we need to scan all transactions from *TID 1* to *TID 6* to get all transactions including item *I1* and *I2* (1、3、5 and 6), then Support(*I1,I2*)=4. For the data in Table 2, we only need to calculate the intersection of TID_sets of *I1* and *I2* and the number of TIDs in the intersection, so Support(*I1,I2*)= Num((1,3,4,5,

* Research Scholar and Assistant Professor, Department of Computer Science, Nehru Memorial College, Puthanampatti, Trichy, Tamilnadu, India.

** Associate Professor, Department of Computer Science, Nehru Memorial College, Puthanampatti, Trichy, Tamilnadu, India

Taxonomy of Dissimilar Stages of Non-Proliferative Diabetic Retinopathy

K.Saraswathi* & Dr.V.Ganesh Babu**

Abstract—Feature selection and extraction is the pre- processing step of Image Mining. Obviously this is a critical step in the entire scenario of Image Mining. Our approach to mine from Images –is to extract patterns and derive knowledge from large collection of images, deals mainly with identification and extraction of unique features for a particular domain. Though there are various features available, the aim is to identify the best features and thereby extract relevant information from the images. Content Based Image Retrieval is the popular image retrieval method by which the target image is to be retrieved based on the useful features of the given image. In this paper, the different Stages of Non-Proliferative Diabetic Retinopathy are identified using the feature extraction from diabetic retina images. The blood vessels in the retina get altered. Exudates are secreted, micro aneurysms and hemorrhages occur in the retina. The appearance of these features represents the degree of severity of the disease. Microaneurysms are focal dilations of retinal capillaries and appear as small round dark red dots. Hemorrhages occur when blood leaks from the damaged retinal vessels. Exudates occur when lipid or fat leaks from abnormal blood vessels. The different stages of Non-Proliferative Diabetic Retinopathy are detected using the features Exudates, Blood Vessels, Microaneurysms and Hemorrhages. An early detection and diagnosis will assist in timely treatment and a reduction in the percentage of visual impairment due to these conditions. it will assist for a better treatment plan and to improve the vision related quality of life.

KEYWORDS—Image Retrieval; Image Mining; Feature Extraction; Retina; Microaneurysms; Hemorrhages; Exudates.

I. INTRODUCTION

Content Based Image Retrieval (CBIR) system prototypes have been proposed and few are used as commercial systems. CBIR aims at searching image databases for specific images that are similar to a given query image. It also focuses at developing new techniques that support effective searching and browsing of large digital image libraries based on automatically derived image features. The features further can be classified as low-level and high-level features. Users can retrieve images based on these features such as texture, color, shape, region and others. By similarity comparison the target image from the image repository is retrieved. The goals of this paper are to discuss about the Different Stages of Non- Proliferative Diabetic Retinopathy using retina features.

II. FEATURE EXTRACTION

Image Mining deals with the extraction of image patterns from a large collection of images. In Image Mining, the goal is the discovery of the image patterns that are significant in a given collection of images [4]. Image mining deals with extraction of knowledge, image data relationship and other required patterns and uses ideas from image processing, image retrieval [1] and machine learning, databases [6].The focus of image mining is on the extraction of knowledge patterns from a large collection of images. While there seems to be some overlap between image mining and content-based retrieval (since both deal with large collection of images), image mining goes beyond the problem of retrieving images. In image mining, the goal is to discover image patterns that are significant in a given collection of images [2]. The different stages of Non-Proliferative Diabetic Retinopathy are detected using the features Exudates, Blood Vessels, Microaneurysms and Hemorrhages.

* Research Scholar,, Bharathiar University, Coimbatore, Assistant Professor in CS, Nehru Memorial College,, Puthanampatti, Trichy (DT), saraswathimuruganmsc@gmail.com

** Assistant Professor and Head, Department of CS, Government College for Women, Maddur, Mandya (DT),Karnataka State, India. vgbzone@gmail.com

PROBLEM SOLVING AND PYTHON PROGRAMMING

Dr. N. DHASARATHAN, K.SARASWATHI,
Dr. R. REKA, Dr. S.R.BOSELIN PRABHU

PROBLEM SOLVING AND PYTHON PROGRAMMING

Dr.N.DHASARATHAN is Working as A Professor of Electronics And Communication Engineering in BVC Engineering College, Amalapuram, East Godavari, A.P. He Is Having 10 Years of Experience in Teaching and Research. He Has Published Around 20 Papers in Various International Journals & Conferences. He Has Authored Two Books Which Includes C Programming & Python Programming.

Mrs.K.SARASWATHI is Working as an Assistant Professor in Nehru Memorial College, Puthanampatti. She has More Than 14 Years of Experience in Teaching and More Than Four Years of Experience in Research. She is Currently Pursuing Her Doctorate Degree at Bharathiar University, Coimbatore. She has Published More Than 21 Papers in Various National, International Journals and Conferences. She Has Written One of The Book Chapters in "Advanced Engineering Research and Applications".

Dr.R.Reka is Working as Professor & Head, Department of Computer Science And Engineering At Annai Mathammal Sheela Engineering College, Erumapatty, Namakkal. She has Above 14 Years of Extensive Working Experience in Teaching and Research. She Got A Gold Medal in master's degree. She has Published Papers More Than 36 in Various National, International Journals & Conferences.

Dr. BOSELIN PRABHU S.R. Obtained his Doctorate (Ph.D) From Anna University Chennai, India. he has 9 Years of Experience in Teaching and Research. he has Published 165 Research Articles in International Journals. he is an Editorial Board Member, Advisory Board Member and Reviewer Of 300 International Journals Both Scopus And SCI Indexed. he Is anElected Fellow Member FUAMAE, FISECE, FISRD, FUAAMP, FISQEM, SIESRP, FUACEE and FISEEE.

Also available as an eBook
ACADEMIC
ISBN 978-93-88719-51-3
9 789388 719513 >

EDUCREATION
PUBLISHING
www.educreation.in

An Enriched framework on Sentiment Analysis of Movie Reviews Data Sets using Machine Learning Algorithm

K.Deepa* & Kirubakaran**

Abstract

In this research paper our target set is movie reviews. There are diverge range of mechanisms to express their data which may be either subjective, objective or a mixture of both. Besides the data collected from World Wide Web consists of lot of movie dataset. It is very much true that we are going to apply some pre-processing techniques and compare the accuracy using Machine Learning algorithm for various classifiers. With ever growing demand to mine the Big Data the open source software technologies.

Keywords: Sentiment Analysis, SVM, K-Nearest Neighbour, Random Forest, Naive Bayes Algorithm.

Introduction

Sentiment analysis refers to the use of natural language processing, text analysis and computational linguistics to extract and identify subjective information in source materials. It

aims to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document [1].

[2]In last decade there is a rise of social media such as blogs and social networks, which has fueled the interest in sentiment analysis. Online opinion has turned into a kind of virtual currency with the proliferation of reviews, ratings, recommendations and other forms of online expression, for businesses that are looking to market their products, identify new opportunities and manage their reputations. In order to automate the process of filtering out

the noise, understanding the conversations, identifying the relevant content and following appropriate actions, many are now looking to the field of sentiment analysis.

[3] The problem of most sentiment analysis algorithms is that they use simple terms to express sentiment about a Product or service. However, cultural factors, sentence negation, sarcasm, terseness, language ambiguity and differing contexts make it extremely difficult to turn a string of written text into a simple pro or con sentiment.

Data Pre-processing

The dataset provided in this competition is comprised of tab-separated files with phrases from the Rotten Tomatoes dataset [4]. The dataset has been divided to training and test set for the purpose of benchmarking, but the sentences have been shuffled from their original order. Each Sentence in the dataset has been parsed into many phrases by the Stanford parser. Each

Phrase has a phrase Id. Each sentence has a sentence Id. Phrases that are repeated (such as short/common words) are only included once in the data.

The train.tsv contains the phrases and their associated sentiment labels. The test.tsv contains Just phrases. Sentiment label to each phrase in test file should be assigned.

* Research Scholar, Bharthidasan University, Tiruchirappalli, India.deepamohan13@gmail.com,Mobile no: 9842129551

** Head, Department of Computer Science, Karunya University, Coimbatore, India.ekirubakaran@gmail.com Mobile no: 9442631150.

Review: Vertical Data Formats Of Frequent Pattern Mining Algorithms

V.Priya*, Dr. S.Murugan**

ABSTRACT: In the current digital world, there is a growth of data for everyday. The researchers are confused by the huge arrival of data. To manage all these massive data is an open challenge for all the researchers in frequent pattern mining. This paper gives a report of the concise narrate of earlier works of vertical formats of frequent patterns with a thorough description of the core algorithms, highlighting the major contributions of the different algorithms put forth by different authors. A broad review of the most significant algorithms of literature has been explained and compared with them in the vertical data Formats.

KEYWORDS: Frequent Pattern Mining, Association Rule Mining, Vertical data formats.

1. INTRODUCTION

Data Mining is the process of discovering knowledge. It is the process of extracting information from available raw data. The data are stored in databases. There are various kinds of data that can be used in data mining which includes transactional data, statistical data etc. Data mining includes various techniques for each purpose. Techniques include Association rule mining, classification, prediction, regression etc. Association rule mining techniques are widely used in discovering hidden correlations and relationship between the set of items in a transaction. It includes every transaction in the database during the discovery process. It also reveals the set of strategies that can be followed or neglected in the field for respective development .Many analysis data mining as synonymous to Knowledge Discovery from Data (KDD), while others consider data mining as an essential stage in the KDD process.

1.1 Data formats of fim algorithms

Frequent item set mining (FIM) is a useful tool for discovering frequently co-concurrent items. Since its inception, a number of significant FIM algorithms have been developed to speed up mining performance. Association rule mining is one of the most important data mining problems. The conviction of association rule mining is the discovery of association relationship among a set of items. There are two different data formats: horizontal data format and vertical data format. The data format of Eclat is vertical data format. Both data formats are the representations of the database in memory. Horizontal data format, a "TID-itemset" format, can also be expressed as "TID: itemset", in which "TID" is the unique identifier of a transaction in a database; "itemset" is the set of items included in the transaction. Vertical data format, an "Item-TID_set" format, can also be expressed as "Item: TID_set", in which "Item" is an item in database, "TID_set" is the set of transactions including the items. Table 1 shows the horizontal data format representation of a database and Table 2 shows the corresponding vertical data format representation. Two data formats have different efficiency in calculating the support degree of the itemset. When adopting the horizontal data format, it needs to scan the whole database to get the support degree of the itemset. But if adopting the vertical data format, it only needs to calculate the intersection of all TID_sets of items to get the support degree. For example, when we want to get the support of an itemset $X=\{I1,I2\}$, for the data in Table 1, we need to scan all transactions from *TID 1* to *TID 6* to get all transactions including item *I1* and *I2* (1、3、5 and 6), then Support(*I1,I2*)=4. For the data in Table 2, we only need to calculate the intersection of TID_sets of *I1* and *I2* and the number of TIDs in the intersection, so Support(*I1,I2*)= Num((1,3,4,5,

* Research Scholar and Assistant Professor, Department of Computer Science, Nehru Memorial College, Puthanampatti, Trichy, Tamilnadu, India.

** Associate Professor, Department of Computer Science, Nehru Memorial College, Puthanampatti, Trichy, Tamilnadu, India

A Survey on Mining Frequent Itemsets

P.Sumathi* & S. Murugan**

Abstract - Data mining is a process of extracting valuable and hidden information from large databases. Association rule mining is one of the popular data mining techniques. In the field of association rules mining, Apriori is most popular algorithm which is used for finding frequent itemsets from large databases. But, the application of Apriori algorithm needs to scan the database many times which leads to a great overhead. In last decades, lots of research activities has been done to improve the efficiency of Apriori algorithm and this paper presents a comprehensive survey of the latest techniques and developments in finding frequent itemset mining and association rule mining.

Keywords - Data Mining, Frequent Itemsets, Association Rule Mining and Apriori Algorithm

Introduction

Association rule mining is one of the popular data mining techniques which have wide range of applications in marketing and retail communities as well as other more diverse fields [6]. Discovery of association rules involves two major steps. 1) Finding frequent itemsets and 2) generating reliable and strong association rules from the frequent itemsets. Step 1 of association rule mining is a challenging task [4] and plays a vital role in mining associations and correlations [5]. Association rules are generated for detecting unknown relationships, producing results which can perform basis for decision making and prediction. Researchers have proposed various improved algorithms for generating frequent itemsets. These algorithms differ in their ways by how they traverse and handle the database. This paper presents an overview of the latest innovations and improvements in finding frequent itemsets proposed by various researchers.

The rest of the paper is organization as follows. In Section 2, describes the Apriori Algorithm and its limitations. In Section 3 elaborates the works related to Apriori and its improvement and the conclusion is given in Section 4.

Apriori Algorithm

Apriori Algorithm is one of the most important algorithms which are used to extract frequent itemsets from large database and obtain the association rule for discovering the knowledge [1] and proposed by R. Agrawal and R. Srikant in 1994 [5]. Apriori employs an iterative approach known as a level-wise search, where k-itemsets are used to explore (k+1)-itemsets. First, the set of frequent 1-itemsets is found by scanning the database to accumulate the count for each item, and collecting those items that satisfy minimum support. The resulting set is denoted by L_1. Next, L_1 is used to find L_2, the set of frequent 2-itemsets, which is used to find L_3, and so on, until no more frequent k-itemsets can be found. The finding of each L_k requires one full scan of the database [15].

To improve the efficiency of the level-wise generation of frequent itemsets, an important property called the Apriori property i.e., all nonempty subsets of a frequent itemset must also be frequent is used to reduce the search space. To generate L_k from L_{k-1}, a two step process is followed. 1) Join step 2)Prune step. Join step generates C_k a set of candidate k-itemsets by joining L_{k-1} with itself and C_k is pruned to generate L_k with "minimum support" as the pruning parameter. To select the best rule among all the available rules, association rule mining depends on two important constraints. They are minimum threshold on support and confidence called minimum support and minimum confidence and is set by domain experts [12]. The support of an association rule is defined as the percentage of the fraction of records that contain to the total number of records in the database.

* Research Scholar, Department of Computer Science, Nehru Memorial College (Autonomous), Puthanampatti, Tiruchirappalli -Dt, TamilNadu, India, 9566385539,sumiparasu@gmail.com

** Associate Professor, Department of Computer Science, Nehru Memorial College (Autonomous), Puthanampatt, Tiruchirappalli -Dt, TamilNadu, India,9486631199,smurugan.nmc@hotmail.com

Praise Worthy Prize

International Review on Computers and Software (I.RE.CO.S.), Vol. 10, N. 6
ISSN 1828-6003
June 2015

Lloyd and Minkowski Based K-Means Clustering for Effective Diagnosis of Heart Disease and Stroke

D. Nalini[1], R. Periasamy[2]

Abstract – *Computer Aided Data mining based decision support system plays a major role in research to easily diagnose medical disease at an early stage. Automatic annotation approach grouped similar medical semantic terms, however, false detection lead to incorrect merging increasing the computational complexity. In this work, a method called Lloyd and Minkowski based K-Means Clustering (LMK-MC) is designed to organize various features in heart disease, with the aim of increasing the features to be incorporated improving the clustering efficiency and reducing the average computational complexity. The Lloyd and Minkowski based K-Means Clustering includes a function that performs a Minkowski based K-Means Clustering to organize similar type of symptom features for easy prediction. By applying Minkowski based K-Means Clustering, similar type of features causing heart disease is analyzed for any set of k centers, aiming at improving the clustering efficiency. Next, the method LMK-MC, applies pair-wise proximities between all pairs of features reducing the computational complexity for intra-cluster. Finally, Lloyd's algorithm based clustering on medical data is designed that moves every center point to the centroid and performs updates. Lloyd's algorithm based clustering obtains local minima solution that easily merges similar features leading to heart disease and strokes, various disease features for labeling is identified effectively with high recognition accuracy. In order to measure the efficiency of LMK-MC, experiments are conducted using Cleveland Clinic Foundation Heart disease data set. Experimental results demonstrate that the proposed Lloyd and Minkowski based K-Means Clustering achieves efficient amount of identification improving the clustering efficiency, reducing the computational complexity while measuring intra-cluster distances for different clusters and, heart disease and stroke identification rate with minimum processing time.*

Keywords: *Lloyd's Algorithm, Minkowski Based K-Means Clustering, Pair-wise Proximities, Centroid*

Nomenclature

PC_{COST}	Pair-wise clustering cost
f_n	*n* features
f_p, f_q	Feature with two instances
C_d	Cluster center
f_d	Feature taken for consideration
k	Total number of groups
M	Matrix
Cl_n	*n* feature of cluster
$Cl_i Cl_{i+1}$	Pair-wise proximities
F_i, F_{i+1}	Features of heart disease

I. Introduction

Computer-aided decision support system based on data mining has received attention from various researchers in order to easily diagnose disease at an early stage. Automatic Annotation Approach using Integrated Interface Schema (AAA-IIS) [1] grouped different medical data with the same semantic aiming at improving the retrieved results.

However, the method increased computational complexity while predicting the diseases as various features were not considered. Another clustering procedure to reduce computational complexity called Warped K-Means (WKM) [2] was introduced. But, with the increase in the features, computational complexity also increased. Due to early prediction of disease to be one of the main solutions in disease diagnosis, in [3], detailed analysis of cardiovascular risk based on classification was presented.

Artificial neural network and feature subset selection was introduced in [4] with the aim of improving the accuracy of disease being predicted over conventional classification techniques.

However, the time with respect to increased features were not considered. To solve this time related issue, non-invasive images methods were introduced in [5] to address the risk related to cardiovascular at an early stage. Another method based on artificial neural network [6] was investigated to improve the detection accuracy.

Though accuracy was increased, the detection rate remained unaddressed.

Secured Communication mechanism in Public Key Cryptography

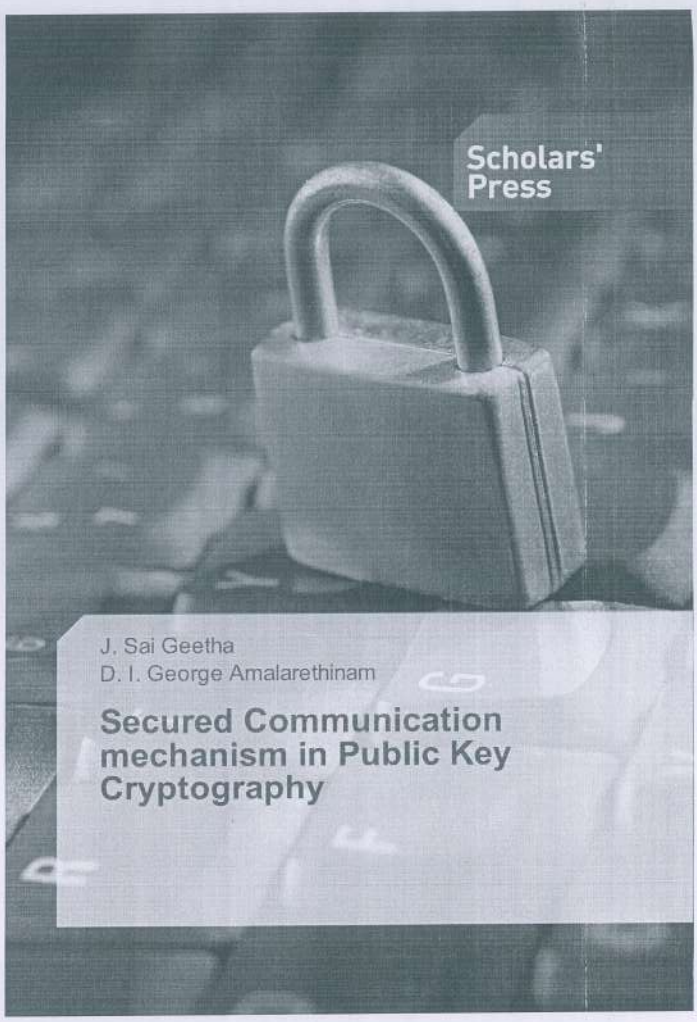
From the ancient days to the information age of today, the volume of information has increased million-fold and acquired economic significance in the context of billions of dollars worth e-commerce. The present socio economic scenario demands speedy transfer of information with enhanced security. In this context, cryptography acquires far greater significance. Cryptography is a method of storing and transmitting data in a secret form which allows correspondence only between the intended persons. The security and speed are obviously two most important parameters in public key cryptosystem. Whenever the security increases, it results in the decrement of speed of crypt process. Hence there is a need for a cryptography algorithm which can optimise both these parameters. In this book,Security is achieved through the proposed algorithms MRGA and ABCRNG. The speed of the algorithm is also enhanced on implementing the algorithms MsgEncA and ModBlockExpoA. The PKCrypt tool has been introduced to support both text and image files. This book will be of useful for researchers working in communication network in the aspect of Security.

Dr.J.Sai Geetha is serving as an Assistant Professor of Computer Science at Nehru Memorial College, Tiruchirappalli,India.
Dr.D.I.George Amalarethinam is Bursar & Director-MCA of Jamal Mohamed College, Tiruchirappalli, India. The authors have several National, International publications and books to their credits.

978-620-2-30632-4

Secured Communication Mechanism in PKC

Sai Geetha, George Amalarethinam



A Hybrid Method Using Rsa with Fast Exponentiation Algorithm

R. Kalaivani*

ABSTRACT— RSA encryption is one of the public-key methods that have been popular in last decade. Considering increment of security requirements, size of the keys has been larger. With key length growing, delay of exponentiation computation has changed into major problem in selecting longer keys. The necessity for information security becomes more widespread in these days, especially for hardware-based implementations such as smart cards chips for wireless applications and cryptographic accelerators. The binary or in other words square-and-multiply method is the classical exponentiation technique that is used in RSA. In this paper a new algorithm of exponentiation in RSA is presented that works in parallel, needs fewer multiplications and so has less delay. Therefore this technique is more useful in larger key computations. Fast modular exponentiation algorithms are often considered of practical significance in public-key cryptosystems.

KEYWORDS--- Cryptography, Decryption, Encryption, Network Security.

I. INTRODUCTION

The RSA cryptosystem is one of the most widely used technologies for achieving information security. The main task of the encryption and decryption engine of RSA cryptosystem is to compute M^E mod N. Because the bit-length of the numbers M, E, and N would be about 512 to 1024 bits now, the computations for RSA cryptosystem are time-consuming. In this paper, an efficient technique for parallel computation of the modular exponentiation is proposed and our algorithm can reduce the time.

II. TYPES OF CRYPTOGRAPHY

Cryptography is one of the most popular techniques that can be used to provide sufficient security to the sensitive data. Cryptography is a method of storing and transmitting data in a particular form so that only those for whom it is intended can read and process it. There are several ways of classifying cryptographic algorithms. Secret Key Cryptography (SKC) uses a single key for both encryption and decryption .Public Key Cryptography (PKC) uses one key for encryption and another for decryption.

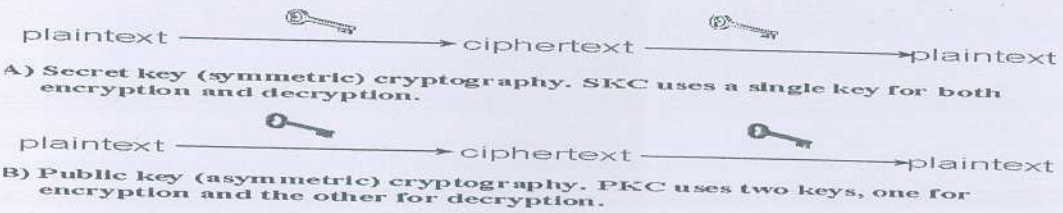


Fig 1: Types of cryptography

III. SYMMETRIC KEY CRYPTOSYSTEM

Symmetric encryption also referred to as conventional encryption or single key encryption was the only type of encryption.

Cipher text = encrypt (plaintext, key) **Plaintext = decrypt (cipher text, key)**

Fig 2 shows the symmetric encryption scheme with five ingredients such as **Plaintext** is the original intelligible message or data that is fed to the algorithm as input.The **Encryption Algorithm** performs various substitutions and permutations on the plaintext. The **Secret Key** is also input to the

* Assistant Professor, Department of CS, Nehru Memorial College, Puthanampatti, Trichy (DT). kalaivinayak@gmail.com

Enhanced Byte Rotation Encryption Algorithm Using Cipher Block Chaining Technique

Velmurugan P*

Abstract

In this digital world protecting the Sensitive data is a required thing. There is a demand for powerful cryptography techniques such as AES, DES, RSA, Blowfish, etc are used to encrypt the data. The quality of the encryption method is measured based on maintaining security and accuracy of the original message. Most of the algorithms achieve data security at greater level, but increasing security can leads to increase time and complexity of the algorithms. This may cause decreasing speed and efficiency of the encryption system. Enhanced Byte Rotation Algorithm (EBREA) using Cipher Block Chaining technique is the new method that ensures speed, security and accuracy of the encryption scheme. Byte Rotation Encryption Algorithm (BREA) works based on Symmetric a key cryptography and Block Cipher algorithm which divides the plain text into n number of plain text blocks using symmetric key technique but it has security issues. To overcome these flaws BREA is enhanced by combining it with Cipher Block Chaining (CBC) algorithm.

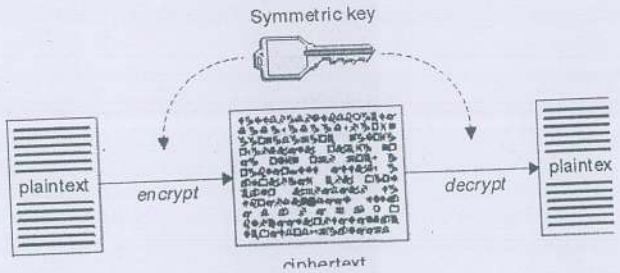
Key Words: Data, Security, Encryption, Decryption, Security, Cipher Block Chaining

I. INTRODUCTION

Cryptography is an art of exchanging the information in the secret manner. It has two main processes they are Encryption and Decryption. Encryption is done by the sender. During the encryption process, the original message (is called as Plain text) is converted into a cipher text by using mathematical operations and some additional information called as key. The generated cipher text has no meaning and no readable form. Then the Cipher text is transferred to the receiver through the network. The receiver receives the cipher text. The receiver decrypts the received cipher text by using any mathematical operations and additional information. This process is called Decryption. At the end of decryption the cipher text is to be converted into original plain text.

II. SYMMETRIC KEY AND ASYMMETRIC KEY CRYPTOGRAPHY

There are two types of crypto systems used to protect the data, commonly known as symmetric key cryptography and asymmetric key cryptography.



Symmetric Key Cryptography

In this cryptography both encryption and decryption process may take same key. This method is also called as private key cryptography.

Asymmetric Key Cryptography

In this cryptography sender and receiver uses different keys to encrypt and decrypt the data. The

* Assistant Professor, Dept of Computer Science, Nehru Memorial College, Puthanampatti, Mobile: 8110982923, Mail: vallivelprof@gmail.com

Taxonomy of Dissimilar Stages of Non-Proliferative Diabetic Retinopathy

K.Saraswathi* & Dr.V.Ganesh Babu**

Abstract—Feature selection and extraction is the pre- processing step of Image Mining. Obviously this is a critical step in the entire scenario of Image Mining. Our approach to mine from Images –is to extract patterns and derive knowledge from large collection of images, deals mainly with identification and extraction of unique features for a particular domain. Though there are various features available, the aim is to identify the best features and thereby extract relevant information from the images. Content Based Image Retrieval is the popular image retrieval method by which the target image is to be retrieved based on the useful features of the given image. In this paper, the different Stages of Non-Proliferative Diabetic Retinopathy are identified using the feature extraction from diabetic retina images. The blood vessels in the retina get altered. Exudates are secreted, micro aneurysms and hemorrhages occur in the retina. The appearance of these features represents the degree of severity of the disease. Microaneurysms are focal dilations of retinal capillaries and appear as small round dark red dots. Hemorrhages occur when blood leaks from the damaged retinal vessels. Exudates occur when lipid or fat leaks from abnormal blood vessels. The different stages of Non-Proliferative Diabetic Retinopathy are detected using the features Exudates, Blood Vessels, Microaneurysms and Hemorrhages. An early detection and diagnosis will assist in timely treatment and a reduction in the percentage of visual impairment due to these conditions, it will assist for a better treatment plan and to improve the vision related quality of life.

KEYWORDS---Image Retrieval; Image Mining; Feature Extraction; Retina; Microaneurysms; Hemorrhages; Exudates.

I. INTRODUCTION

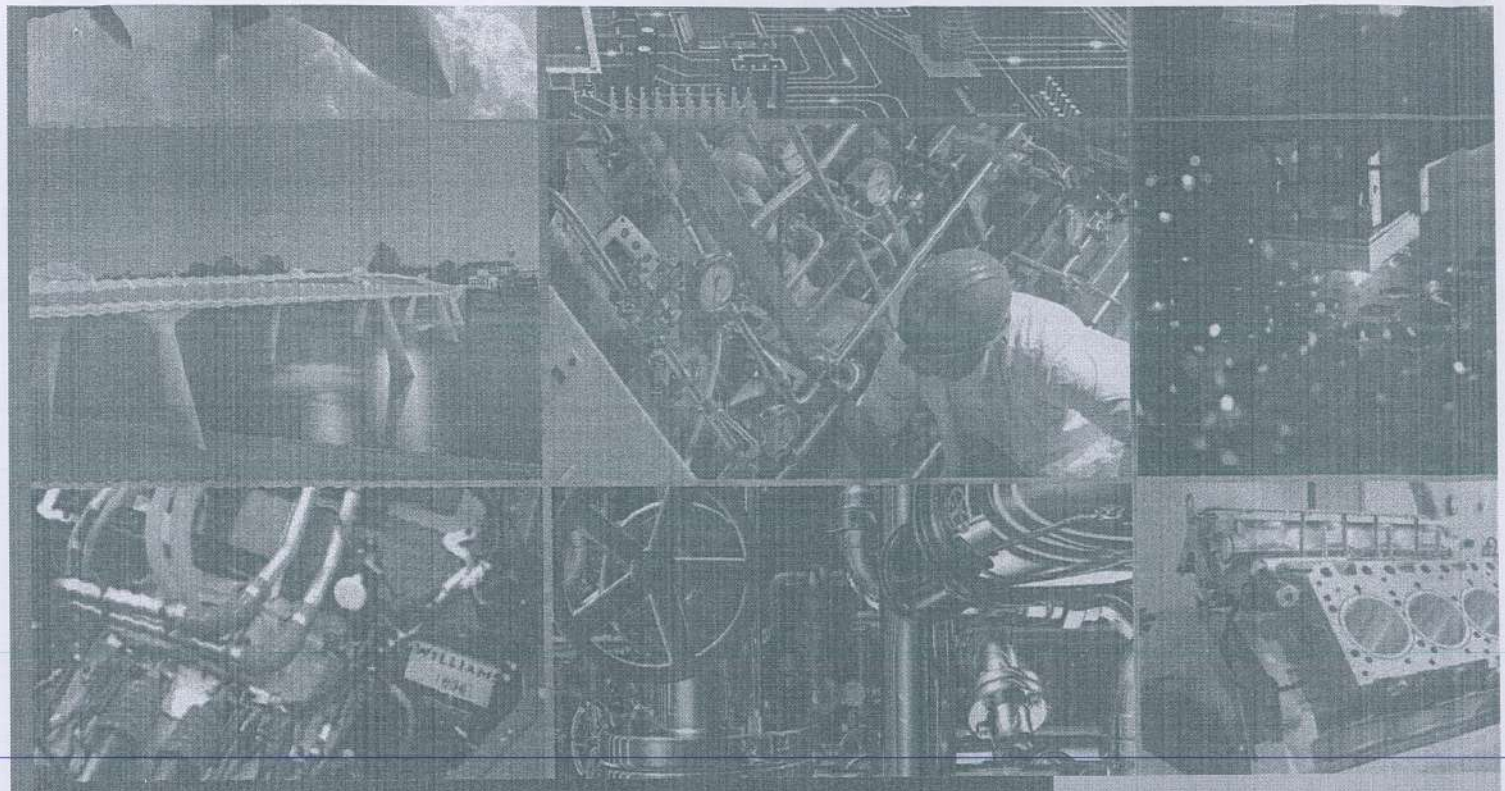
Content Based Image Retrieval (CBIR) system prototypes have been proposed and few are used as commercial systems. CBIR aims at searching image databases for specific images that are similar to a given query image. It also focuses at developing new techniques that support effective searching and browsing of large digital image libraries based on automatically derived image features. The features further can be classified as low-level and high-level features. Users can retrieve images based on these features such as texture, color, shape, region and others. By similarity comparison the target image from the image repository is retrieved. The goals of this paper are to discuss about the Different Stages of Non- Proliferative Diabetic Retinopathy using retina features.

II. FEATURE EXTRACTION

Image Mining deals with the extraction of image patterns from a large collection of images. In Image Mining, the goal is the discovery of the image patterns that are significant in a given collection of images [4]. Image mining deals with extraction of knowledge, image data relationship and other required patterns and uses ideas from image processing, image retrieval [1] and machine learning, databases [6].The focus of image mining is on the extraction of knowledge patterns from a large collection of images. While there seems to be some overlap between image mining and content-based retrieval (since both deal with large collection of images), image mining goes beyond the problem of retrieving images. In image mining, the goal is to discover image patterns that are significant in a given collection of images [2]. The different stages of Non-Proliferative Diabetic Retinopathy are detected using the features Exudates, Blood Vessels, Microaneurysms and Hemorrhages.

* Research Scholar,, Bharathiar University, Coimbatore, Assistant Professor in CS, Nehru Memorial College,, Puthanampatti, Trichy (DT), saraswathimuruganmsc@gmail.com

** Assistant Professor and Head, Department of CS, Government College for Women, Maddur, Mandya (DT),Karnataka State, India. vgbzone@gmail.com



Volume - II

Advanced Engineering Research and Applications

Editor-in-Chief:
Hongseok Choi

Research India Publications

CHAPTER - 18

Entropy Based Image Retrieval Method for Retrieving the images and Feature Based Image Recovery Method for Extracting the Features via Image Gathering and Reclamation using Image Mining Procedures

K.Saraswathi[1]
[1]*Assistant Professor (Research Scholar)*
Department of CS, Nehru Memorial College, Puthanampatti, Trichy (DT), India.
E-mail: saraswathimuruganmsc@gmail.com

Dr.V.Ganesh Babu[2]
[2]*Assistant Professor, Department of CS*
Government College for Women, Maddur, Mandya (DT), Karnataka State, India.
E-mail: vgbzone@gmail.com

Abstract

Presently, tools for mining images are few and require human interventions. Feature selection and extraction is the pre-processing step of Image Mining. Obviously this is a critical step in the entire scenario of Image Mining. Our approach to mine from Images –is to extract patterns and derive knowledge from large collection of images, deals mainly with identification and extraction of unique features for a particular domain. Though there are various features available, the aim is to identify the best features and thereby extract relevant information from the images. Various methods for extraction are used in this paper. Content Based Image Retrieval is the popular image retrieval method by which the target image is to be retrieved based on the useful features of the given

Price: Within India: Rs. 1500
Outside India: US$ 125

ISBN: 978-93-86138-58-3

RIP

Research India Publications
Head Office: B-2/84, Ground Floor,
Rohini Sector-16, Delhi-110089, INDIA
Fax No.: +91-11-27297815
Email: ripublication@vsnl.net
Website: www.ripublication.com

Image Encryption and Decryption in Public Key Cryptography based on MR

D.I.George Amalarethinam
Associate Professor & Director- MCA,
Jamal Mohamed College
Trichy, India
di_george@ymail.com

J.Sai Geetha
Assistant Professor, Dept of Computer Science
Nehru Memorial College, Puthanampatti
Trichy, India
jsaigeetha99@gmail.com

Abstract— In the past decade, image encryption is given much attention in research of information security and a lot of image encryption algorithms have been introduced. Due to some intrinsic features of images like bulk data capacity and high data redundancy, the encryption of image is different from that of text; therefore it is difficult to handle them by traditional encryption methods. In the proposed work, a new image encryption algorithm based on Magic Rectangle (MR) is being applied. To begin with, the plain-image is converted into blocks of single bytes and then the block is replaced as the value of MR. Further, the control parameters of Magic Rectangle (MR) are selected randomly by the user. Subsequently the image is being encrypted with public key cryptography algorithms such as RSA, ElGamal etc. The experimental result shows that the proposed algorithm can successfully encrypt/decrypt the images with separate secret keys, and the algorithm has good encryption effect. Cipher text developed by this method will be entirely different when compared to the original image file and will be suitable for the secured transmission over the internet. Thus, this model provides an additional level of security to public key algorithm and efficient utilization of memory.

Keywords— Communication Security, Image Encryption and Decryption, Public Key Cryptography, Public key, Secret key, Magic Rectangle.

I. INTRODUCTION

Computer has become an essential device now a days. The main use of computer is to store data and send it from one location to other. The information that is shared must be transferred in a secured manner. To ensure secured transmission of information, data is encrypted to unreadable formats by an unauthorized person. Cryptography is the science of information security which has become a very critical aspect of modern computing systems towards secured data transmission and storage. The exchange of digital data in cryptography results in different algorithms that can be classified into two cryptographic mechanisms: symmetric key in which same key is used for encryption and decryption and asymmetric key in which different keys are used for encryption and decryption [1]. Asymmetric key algorithms are more secured when compared with symmetric key algorithms.

Nowadays, information security is primarily based on data storage and transmission. Images are broadly used in numerous processes. As a result, the safety of image data from unauthorized access is crucial at the hands of user. Image encryption plays a significant role in the field of information hiding. Image hiding or encryption methods and algorithms ranges from simple spatial domain methods to more complicated and reliable frequency domain.

The parameters used in encryption and decryption process of the algorithm plays a vital role for security such as key streams in one time pad, the secret key in Data Encryption Standard algorithm, the prime p and q in RSA etc. Of all the encryption algorithms available, RSA(Rivest, Shamir, Adlemen) accounts for highly reliable one. In RSA, the secret key is derived from the public key and choosing p and q with very large size. Even though the above parameters are considered, it is not fully secured. The conventional method of image encryption is done through any one of the technique such as RGB color shuffling, bits manipulation, chaotic mapping method etc. Of all the methods stated above, the result of encryption is in the form of cipher image. It takes more time for encryption and decryption process and the inefficient use of memory in this kind of cryptosystem results in reduction of transmission speed. To overcome this problem, this paper tries to develop a entirely different method by introducing a special singly even magic rectangle [2] of the order 32 x 48. Thus preferably different numerals representing the bytes of image values are taken from magic rectangle instead of taking patterns or bits for encryption. The encryption process is being performed using RSA cryptosystem.

II. RELATED WORK

Quist-Aphetsi Kester[3], proposed the work sets out to contribute to the general body of knowledge in the area of cryptography application and by developing a cipher algorithm for image encryption of m*n size by shuffling the RGB pixel values. Finally, the algorithm made it possible for encryption and decryption of the images based on the RGB pixel.

Musheer Ahmad and M. Shamsher Alam [4] proposed a new image encryption algorithm based on three different chaotic maps. In this work, the plain-image is first decomposed into 8x8 size blocks and then the block based shuffling of image is

ISBN:978-1-4799-7623-2/15/$31.00 ©2015 IEEE

International Journal of Pure and Applied Mathematics
Volume 117 No. 9 2017, 125-128
ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
doi: 10.12732/ijpam.v117i9.22
Special Issue

RPC Approach to Know the Level of Microaneurysms in Mild Juncture of Nonproliferative Diabetic Retinopathy

K.Saraswathi [1], Dr.V.Ganesh Babu[2], Dr.R.Reka[3],Dr.T.Parithimarkalaignan[4]

[1]Research Scholar, Bharathiar University, Assistant Professor, Nehru Memorial College, Puthanampatti, Trichy (DT), Tamilnadu, India.
[2]Assistant Professor & Head of the Department,Department of Computer Science, Government College for Women, Maddur, Mandya (DT), Karnataka, India.
[3]Professor & Head of the Department, Department of Computer Science and Engineering, Annai Mathammal Sheela Engineering College, Erumapatti, Namakkal (DT), Tamilnadu, India.
[4]The Principal, Annai Mathammal Sheela Engineering College, Erumapatti, Namakkal (DT), Tamilnadu, India.

rekasatheesh@gmail.com[1] saraswathimuruganmsc@gmail.com[2] vgbzone@gmail.com[3]

Abstract
Prolongation of Nonproliferative Diabetic Retinopathy may result in permanent blindness. A group of eye conditions that affect people with diabetes which causes progressive damage to the tiny blood vessels in the retina. We scrutinized the retina to know the mild level of Nonproliferative Diabetic Retinopathy through the count of red pixels in Microaneurysms image. The Mild stage of NPDR is detected through Microaneurysms. Exposure of Nonproliferative Diabetic Retinopathy at early stage is recommended. The blood vessels in the retina get altered, secretion of exudates, micro aneurysms and hemorrhages occur in the retinas. The appearance of these features represents the degree of severity of the disease. Microaneurysms are focal dilations of retinal capillaries and appear as small round dark red dots .The Microaneurysms appear as small blood clots on the surface of the retina layers. The Mild stage is indicated by the presence of at least 1 Microaneurysms on the surface of the retina layers. The level of microaneurysms in mild juncture of nonproliferative diabetic retinopathy is identified through the count of red pixels in the retina.
Keywords: Retina, Microaneurysms, Red Pixels.

1 Introduction

The level of Microaneurysms is detected through the count of red pixels in the retina. Then the same process is applied in retina with Microaneurysms. The mild juncture of Nonproliferative Diabetic Retinopathy is detected through the count of red pixels. The retina is easily detected whether it is affected by Diabetic Retinopathy or not using the red pixels count.

1.1 Methods

The retina image is converted into green channel image. Adaptive histogram equalization technique is adopted to perform the Contrast enhancement. Then Morphological filling is performed on the green channel image. The unfilled green channel image is then subtracted from the filled one. The RGB color image is converted into binary image. Then the background and foreground pixels are changed to yield an image with Microaneurysms patches. The levels of Microaneurysms are detected through the count of red pixels.

2 Methods of Detection

2.1 Fundus Image

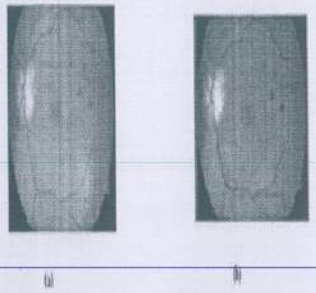
Images of retina are taken by a device called fundus camera. This camera takes images of the internal surface of retina, posterior pole, macula, optic disc, and blood vessels. Some benchmark databases are openly available for the assessment of algorithms introduced for the computerized screening and analysis of DR. The purpose of these databases is to check the strength of automatic screening of DR and then compare the results with current techniques. Seven datasets are available openly including DRIVE, STARE, DIARETDB, E-ophtha, HEI-MED, Retinopathy Online Challenge (ROC), and Messidor [6].We took the images from STARE datasets. In Fig.1 (a) represented original retina fundus image taken from STARE dataset.

2.2 Morphological Filling

The original image is converted into green channel image. In Fig.1 (a) represented original image and (b) represented green channel image. The contrast of the grayscale image is enhanced by transforming the values using contrast-limited adaptive histogram equalization (CLAHE) [3]. Histogram equalization redistributes the histogram of each color channel in the input image such that the output image contains a uniform pixel value distribution. The assumption is that for each color plane the pixel rank order is maintained even with variations in illumination. A monotonic, non-linear transformation function is applied to equalize the histogram of each separate color channel [9]. An output image is produced by applying the function to map each gray level value in the input image to a new corresponding value in the output image [4].In Fig1 (c) represented histogram equalization image. The structuring element function created STRELs of any arbitrary size and shape. Here this function created square shaped two STRELs objects. Then the morphological close operation is performed on the output image which joined the individual objects in the output image together by filling in the gaps between them and by smoothing their outer edges [10].The holes in the output image is filled by imfill function. In Fig1 (d) represented the output image after morphological operations.

2.3 Microaneurysms Image and Red Pixels Count in Mild Juncture

The output image generated by imclose operation is subtracted from morphological filling image which is represented by a Fig. 2(a).The subtracted image is converted into binary image through the intensity value in the subtracted image. The noise in the original image is removed through median filter and preserved edges [8].The constant value is subtracted from original image to receive the marker image. The image after noise removal is reconstructed using marker image. The background image is retrieved through the subtraction of marker image from original image. Then the binary image is received through the intensity value of the background image. The components are connected in the image. The new binary image is received after the removal of objects that have fewer than some constant pixels (p) from old binary image. The new binary image is represented by Fig.2 (b).The new binary image is subtracted from old binary image. Then the vessels are extracted from the retina by using some constant threshold value in histogram equalization image. The eroded image is received through the binary erosion. Then the each element in the eroded image is multiplied with each element in the original image. The morphological operation is performed in the output image. The image after morphological operation is converted into binary image. Microaneurysms may appear in isolation or in clusters as tiny, dark red spots. Their sizes range from 10-100 microns i.e. less than 1/12th the diameter of an average optics disc and are circular in shape [7].From analysis and experiment, the pixel count for candidate microaneurysms ranges from 30 to 5000 pixels for a (1320x1024) image. An area less than the range of 30 to 5000 pixels are regarded as a background noise. Then the microaneurysms image is received. Then the binary image [I_micro] is converted into RGB image. In this image the microaneurysms are displayed in red color which is represented by Fig.2(c).The red pixels are counted and displayed as red pixels count in Table1.



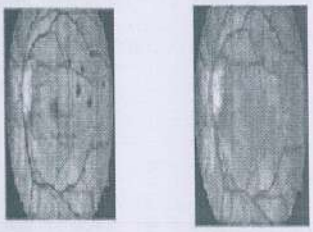


Fig.1 (a) Original Image (b) Green Channel Image
(c) Histogram Equalization Image (d) Image after Morphological Filling

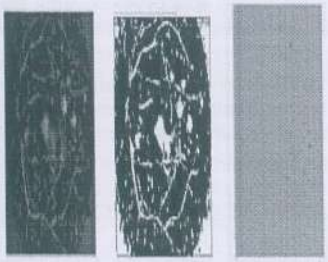


Fig.2 (a) Subtraction Image (b) Binary Image (c) Microaneurysm Image

2.4 Level of Microaneurysms in Mild Juncture

The number of red pixels are increased whenever the number of microaneurysms are increased which are shown in Table1.The count of red pixels for different retinas are plotted in Fig. 3.

S.No	Image Names	Retina Image	Result Image	Red Pixels Count
1.	Image1			69037
2.	Image2			64138
3.	Image3			69343
4.	Image4			68090
5.	Image5			69799
6.	Image6			66725
7.	Image7			719299
8.	Image8			681244
9.	Image9			660254
10.	Image10			66572

Table1. Red Pixels Count for Different Retinas in Mild Juncture of NPDR

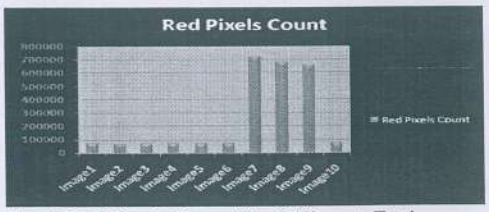


Fig.3 Red Pixels Count for Different Retinas Dissertation

Conclusion

Eye deceases like Diabetic retinopathy (DR) is responsible for blindness in human eye. Therefore it is necessary to detect such deceases at early stage with the help of image processing technologies and methods. RPC is able to know the level of microaneurysms in different retinas in mild juncture of non proliferative diabetic retinopathy. In this paper we discussed about the variation of red pixels count in mild juncture of Nonproliferative diabetic retinopathy. So we can conclude the eye is affected by mild or heavy through the red pixels count in the retina.

Acknowledgements

I am thankful to Dr. V.Ganesh Babu (Assistant Professor in Computer Science, Government College for Women, Maddur, Mandya, and Karnataka) for his excellent guidance and K.Vasantha (Mother), N.Karuppanna (Father) and T.Murugan (Spouse) for supporting me in all the ways.

References

[1] Jyoti D. Patil., Anant. L. Chaudhari., Tool for the Detection of Diabetic Retinopathy using Image Enhancement Method in DIP. Int. J. Applied Information Systems (IJAIS) – *ISSN*: 2249-0868, FCS, New York, USA.

[2] Aniruddha L., Dr.Srikanth Prabhu., Dr.Niranjana Sampathila., Detection of Abnormal Features in Digital Fundus Image Using Morphological Approach for Classification of Diabetic Retinopathy. Int. J. Innovative Research in Computer and Communication Engineering – ISSN: 2320-9801,Vol. 3, Issue 2(2015)

[3] Carmen Valverde., Maria Garcia1., Roberto Hornero1., Maria I Lopez-Galvez2., Automated detection of diabetic retinopathy in retinal images. http://www.ijo.in.

[4] Pavle Prentasi., Detection of Diabetic Retinopathy in Fundus Photographs. Unska 3, 10000 Zagreb, Croatia.

[5] Falguni Thakkar., Rajvi Parikh., A Survey on Automatic Detection of Diabetic Retinopathy Exudates from Retinal Fundus Images. Int. J. Advanced Research in Computer and Communication Engineering-ISSN: 2278-1021, Vol. 5, Issue 5(2016).

[6] Javeria Amin., Muhammad Sharif., and Mussarat Yasmin., A Review on Recent Developments for Detection of Diabetic Retinopathy. Scientifica. Article ID 6838976(2016).

[7] Zahira Asifa Tarannum., B.Srilatha. Detection of Diabetic Retinopathy with Feature Extraction using Image Processing.Int.J. Electrical, Electronics and Computer Systems.ISSN: 2347-2820, Vol .3, Issue-8(2015).

[8] M. Jagannath., K. Adalarasu., Diagnosis Of Diabetic Retinopathy From Fundus Image Using Fuzzy C-Means Clustering Algorithm. IIOAB Journal.ISSN: 0976-3104, Vol.6, Issue 4(2015).

[9] N.S.Datta., R.Sarker. H.S.Dutta., M.De. Software based Automated Early Detection of Diabetic Retinopathy on Non Dilated Retinal Image through Mathematical Morphological Process. Int.J.Computer Applications (0975 – 8887). Vol. 60, No.18 (2012).

[10] A.S.Jadhav., Pushpa., B.Patil., Classification Of Diabetes Retina Images Using Blood Vessel Area. Int. J.Cybernetics & Informatics. Vol. 4, No. 2(2015).

Xiwei Zhang (2014) Image processing methods for computer-aided screening of diabetic retinopathy. Dissertation, MINES ParisTech.