Visvesvaraya Technological University
Belgaum-590014

SEMINAR REPORT

On

"AN ARTIFICIAL INTELLIGENCE DRIVEN MULTI FEATURE EXTRACTION SCHEME"

Submitted in partial fulfilment of the requirements for the award of the degree of ***Bachelor of Engineering in Computer Science and Engineering*** *of Visvesvaraya Technological University, Belgaum.*

Submitted by:

KIRAN KUMAR H (1AM16CS076)

Under the Guidance of

Dr. LATHA C.A

Prof., & Head, Dept. of CSE

Department of Computer Science and Engineering

AMC Engineering College

(NAAC & NBA Accredited, Approved by AICTE, New Delhi & Affiliated to VTU, Belagavi)

18th K.M, Bannerghatta Main Road, Bangalore-560 083

2019-2020

(Accredited by NAAC & NBA, Ministry of Hrd, New Delhi & Affiliated to VTU Belgavi)

PRINCIPAL
AMC ENGINEERING COLLEGE
BENGALURU - 560 083

AMC ENGINEERING COLLEGE

(NAAC & NBA ACCREDITED, APPROVED BY AICTE, NEW DELHI & AFFILIATED TO VTU, BELAGAVI)
BENGALURU– 560083

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

CERTIFICATE

This is to certify that the internship work entitled **"AN ARTIFICIAL INTELLIGENCE DRIVEN MULTI FEATURE EXTRACTION SCHEME"** has been successfully carried out by **KIRAN KUMAR H (1AM16CS076)** student of **AMC Engineering College** in partial fulfilment of the requirements for the award of degree in **Bachelor of Engineering in Computer Science and Engineering** of **Visvesvaraya Technological University, Belgaum** during academic year 2019-2020. It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the report deposited in the departmental library.

Guide & HOD:
Dr. LATHA C.A
Prof., & Head, Dept. of CSE

Principal:
Dr. A.G NATARAJ
Principal

Examiners: **Signature with Date:**

1. _____ _____

2. _____ _____

Guide

PRINCIPAL
AMC ENGINEERING COLLEGE
BENGALURU - 560 083.

DECLARATION

I the undersigned student of 8th semester Department of Computer Science & Engineering, AMC Engineering College, declare that my internship work entitled "IoT Based Vehicular Management and Diagnostics System" is a bonafide work of Mine. My internship is neither a copy nor by means a modification of any other engineering project.

I also declare that this project was not entitled for submission to any other university in the past and shall remain the only submission made and will not be submitted by me to any other university in the future.

Name	USN	Signature
1. _____	_____	_____

PRINCIPAL
AMC ENGINEERING COLLEGE
BENGALURU - 560 083.

ABSTRACT

The Internet improves the speed of information dissemination, and the scale of unstructured text data is expanding and increasingly being used for mass communication. Although these large amounts of data meet the infinite demand, it is difficult to find public focus in a timely manner. Therefore, information extraction from big data has become an important research issue, and there are many published studies on big data processing at home and abroad. In this paper, we propose a multi-feature keyword extraction method, and based on this, an artificial intelligence driven big data MFE scheme is designed, then an application example of the general scheme is expanded and detailed. Taking news as the carrier, this scheme is applied to the algorithm design of hot event detection. As a result, a multi-feature fusion clustering algorithm is proposed based on user attention with two main stages. In the first stage, a multi-feature fusion model is developed to evaluate keywords, and this model combines the term frequency and part of speech features. We use it to extract keywords for representing news and events. In the second stage, we perform clustering and detect hot events in accordance with the procedure, and during the composition of news clusters, we analyze several variadic parameters in order to explore the optimal effectiveness. Then, experiments on the news corpus are conducted, and the results show that the approach presented herein performs well.

ACKNOWLEDGEMENT

First and foremost, I would like to thank GOD, the Almighty for being so merciful on me. He guided me in every walk of life to do something good and hence this dissertation.

I have a great pleasure in expressing my deep sense of gratitude to founder **Chairman Dr. K. R. Paramahamsa** for having provided me with a great infrastructure and well-furnished labs.

I express my sincere thanks and gratitude to our **Principal Dr. A.G. Nataraj** for providing me an opportunity to carry out my dissertation work.

I would like to extend my special thanks to **Prof. Latha C A, HOD, Department of** CSE for her support and encouragement.

I am grateful to my guide **, Prof. Latha C A, HOD, Department of CSE, AMC Engineering College, Bangalore** for her unfailing encouragement and suggestion, given to me in the course of my dissertation work.

I am also grateful to all the staff members of the Department of Computer Science & Engineering for their encouragement and support.

Last but not the least, I wish to thank all my friends and family members for their help and co-operation.

Place: Bengaluru
Date:

Kiran Kumar H
(1AM16CS076)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATION

AAP	Annular Accumulated Points
BDCT	Block Discrete Cosine Transform
CCPM	Conditional Co-occurrence Probability Matrix
CNN	Convolutional Neural Network
DCT	Discrete Cosine Transform
JPEG	Joint Photographic Experts Group
PCA	Principal Component Analysis
SVM	Support Vector Machine

CHAPTER 1

INTRODUCTION

In this chapter there's a discussion on the basic concepts of artificial intelligence, importance of big data in our digital world, various types of data, basic foundation of information extraction and text mining concepts.

1.1 Artificial Intelligence

Artificial intelligence (AI) is wide-ranging branch of computer science concerned with building smart machines capable of performing tasks that typically require human intelligence. AI is an interdisciplinary science with multiple approaches, but advancements in machine learning and deep learning are creating a paradigm shift in virtually every sector of the tech industry. In computer science, artificial intelligence (AI), sometimes called machine intelligence, is intelligence demonstrated by machines, AI is a computer system able to perform tasks that ordinarily require human intelligence. Many of these artificial intelligence systems are powered by machine learning, some of them are powered by deep learning and some of them are powered by very structured rules. Turing's paper "Computing machinery and Intelligence" (1950), and its subsequent Turing Test, established the fundamental goal and vision of artificial intelligence. As machines become increasingly capable, tasks considered to require "intelligence" are often removed from the definition of AI, a phenomenon known as the AI effect. In the twenty-first century, AI techniques have experienced a high rise in the following concurrent advances in computing power, large amounts of data, and theoretical understanding; and AI techniques have become an essential part of the technology industry, helping to solve many challenging problems in computer science, software engineering and operation research.

Many problems in AI can be solved in theory by intelligently searching through many possible solutions, Reasoning can be reduced to performing a search. For example, logical proof can be viewed as searching for a path that leads from premises to conclusions, where each step is the application of an inference rule. Planning algorithms search through trees of goals and sub goals, attempting to find a path to a target goal, a process called means-ends analysis. Robotics algorithms for moving limbs and grasping objects use local searches in configuration space. Many

learning algorithms use search algorithms based on optimization. AI automates repetitive learning and discovery through data, AI adapts through progressive learning algorithms.

1.2 Impact of Big Data

Big data is a field that treats ways to analyze, systematically extract information from, or otherwise deal with data sets that are too large or complex to be dealt with by traditional data-processing application software. Data with many cases (rows) offer greater statistical power, while data with higher complexity (more attributes or columns) may lead to a higher false discovery rate. Big data challenges include capturing data, data storage, data analysis, search, sharing, transfer, visualization, querying, updating, information privacy and data source. Big data was originally associated with three key concepts: volume, variety, and velocity. When we handle big data, we may not sample but simply observe and track what happens. Therefore, big data often includes data with sizes that exceed the capacity of traditional software to process within an acceptable time and value. Data sets grow rapidly, to a certain extent because they are increasingly gathered by cheap and numerous information-sensing Internet of things devices such as mobile devices, aerial (remote sensing), software logs, cameras, microphones, radio-frequency identification (RFID) readers and wireless sensor networks. The world's technological per-capita capacity to store information has roughly doubled every 40 months since the 1980s; as of 2012, every day 2.5 exabytes (2.5×2^{60} bytes) of data are generated. Based on an IDC report prediction, the global data volume will grow exponentially from 4.4 zettabytes to 44 zettabytes between 2013 and 2020. By 2025, IDC predicts there will be 163 zettabytes of data. One question for large enterprises is determining who should own big-data initiatives that affect the entire organization.

1.3 Information Extraction

Information extraction (IE) is the task of automatically extracting structured information from unstructured and/or semi-structured machine-readable documents. In most of the cases this activity concerns processing human language texts by means of natural language processing (NLP). Recent activities in multimedia document processing like automatic annotation and content extraction out of images/audio/video/documents could be seen as information extraction. Information Extraction

is the part of a greater puzzle which deals with the problem of devising automatic methods for text management, beyond its transmission, storage and display. The discipline of information retrieval (IR) has developed automatic methods, typically of a statistical flavor, for indexing large document collections and classifying documents. Another complementary approach is that of natural language processing (NLP) which has solved the problem of modelling human language processing with considerable success when taking into account the magnitude of the task.

Information extraction depends on named entity recognition (NER), a sub-tool used to find targeted information to extract. NER recognizes entities first as one of several categories such as location, persons or organizations (ORG. Once the information category is recognized, an information extraction utility extracts the named entity's related information and constructs a machine-readable document from it, which algorithms can further process to extract meaning. IE finds meaning by way of other subtasks including co-reference resolution, relationship extraction, language and vocabulary analysis and sometimes audio extraction.

There are two types of data present in the form of text in internet, these are structured text form and unstructured form of data, Unstructured data (or unstructured information) is information that either does not have a pre-defined data model or is not organized in a pre-defined manner. Unstructured information is typically text-heavy, but may contain data such as dates, numbers, and facts as well. This results in irregularities and ambiguities that make it difficult to understand using traditional programs as compared to data stored in fielded form in databases or annotated (semantically tagged) in documents.

A structured text is an abstract model that organizes elements of data and standardizes how they relate to one another and to the properties of real-world entities. For instance, a data model may specify that the data element representing a car be composed of a number of other elements which, in turn, represent the color and size of the car and define its owner. The term data model can refer to two distinct but closely related concepts. Sometimes it refers to an abstract formalization of the objects and relationships found in a particular application domain.

1.4 Introduction about the proposed system

In this work we propose a multi-feature keyword extraction (MFE) for calculating the relevance of words and phrases to the content of the article subject and returning the first N words or phrases that can best represent the article subject according to the order of relevance. The evaluation method of multi-feature fusion is used in the process of keyword extraction, which can extract high-quality keywords even if the article is short. Then combining with the features of user attention, such as article reading quantity, comment volume and comment growth rate, we adopted a new algorithm based on MFE to cluster text of various media, so as to facilitate subsequent analysis of social hot events. We'll take an example of online news as a major example in our work. Online news reports have increased considerably among enormous quantities of data on blogs, online newspapers and news websites. The massive amount of news can overwhelm people when they access reports of interest, although there are currently general classifications of news, such as business, technology, and sports. For an individual who frequently reads news, it would be desirable if they could access or abandon all similar news reports conveniently targeting events in which they are personally interested or uninterested. To accomplish this, the clustering model is built. Usually, a news report includes not only the contents of the report itself but also some incidental information, such as the number of comments, which is often overlooked. In addition to concentrating on various things.

we expect incidental information to play its due role. Depending on this case, the following two aspects will be taken into consideration. One is an improved representation of news reports, which should effectively model the contents. The other is that hot news can better represent events, so we propose a method based on user attention for event extraction.

According to this case, our main contributions include the following:

1. Based on tens of thousands of online news from NetEase News, we analyzed the characteristics of static pool with news to extract events that most people focus on.
2. We proposed a new method to calculate the similarity between news and events, relying on an aligned set of basis vectors obtained using keyword correlation analysis.
3. We established a multi-feature fusion model for evaluating keywords that combines the term frequency and part of speech features together.

4. We designed an approach to detect hot events in accordance with the procedure, and during the composition of news clusters, we analyzed several variadic parameters in order to explore the optimal effectiveness.

CHAPTER 2

LITERATURE SURVEY

A Literature survey describes how this concept has emerged, how it has been implemented and what is the current status. This research mainly relates to previous work on information extraction, news event detection and keyword extraction. Thus, we mainly review them in this chapter.

During the last few years, information extraction has attracted wide attention due to adding knowledge to the search results of major commercial search engines. Working with collections of articles in a particular domain to extract relevant information in a structured manner are often customized. This means some specific templates are used for filling in with information collected from texts. The main obstacle is poor portability because templates are designed for a specific purpose and they are difficult to reuse.

In Paper [1], the author describes an information extraction algorithm for identifying entities and relationships in texts. Info boxes and Wikipedia are used to solve the problem of named entity recognition in information extraction, this procedure might face a problem of mixed structured texts, the solution for this would be a method for identifying terms in articles that are proposed, using these terms to identify document-related fragments. In the paper [2] they describe a method describes how to summarize a web entity based on the entity's occurring in web articles. We can refer to [3] for the knowledge generation. It describes a project that attempts to automatically extract information about artists from the Web, use it to generate personalized narrative biographies, and populate a knowledge base.

YAGO is well-known project [3] related to extracting knowledge from WordNet and Wikipedia. It presents an efficient technique to extract immediate information withing very less time. An automatic query-based retrieval method [4] has been developed to extract user-defined relationships from large text databases such as media databases, audio databases, and text databases, using these databases we can find a relationship within this data to get some insights. This method is used on structured type of data where we have the data in the form of tables. Automatic query-based system is used to give some keywords and retrieve any information based on our inputs.

In paper [5] it contributes to the computation of objective evaluations with a tool that is targeted at extracting data about companies from web accessible, semi-structured resources in a way that fits the justly tight requirements of the platform. This tool has an efficient functionality to gather as much data as possible about the companies from the semi structured way. This fits into the correct way and accomplishes the requirements that has to be met for the information extraction. In paper [6] this describes that machine learning has gained much more interest due to effectiveness and efficiency, particularly their success in many shared tasks. Some machine learning methods were directly used for natural language processing task such as tumor information extraction where we try to gather the information whether the patient has tumor or not based on past information of the patient. There are various types of tumors like benign and malignant where the final results show the probability what type of tumor the patient has been suffering.

In paper [7] Information extraction is a hot topic in the field of natural language processing in recent years, in which event extraction is one of the three main tasks. The ACE (Automatic Content Extraction) evaluation conference, Automatic content extraction (ACE) is a research program for developing advanced information extraction technologies convened by the from 1999 to 2008, succeeding MUC and preceding Text Analysis Conference which promotes the development of event extraction, defines events as special things that involve participants with unique ways to improve the extraction schemes reduce the time limit for getting the information.

The ACE program, however, defines the research objectives in terms of the target objects (i.e., the entities, the relations, and the events) rather than in terms of the words in the text. For example, the so-called "named entity" task, as defined in MUC, is to identify those words (on the page) that are names of entities. In ACE, on the other hand, the corresponding task is to identify the entity so named. This is a different task, one that is more abstract and that involves inference more explicitly in producing an answer.

The Message Understanding Conferences (MUC) were initiated and financed by DARPA (Defense Advanced Research Projects Agency) to encourage the development of new and better methods of information extraction. The character of this competition is concurrent research teams competing against one another that is required for the development of standards for evaluation,

e.g. the adoption of metrics like precision and recall. In the MUC (Message Understanding Conference) evaluation meeting before the ACE meeting, there is a scenario template task, which focuses on the extraction of events. As early as 1958, a classical approach was based on the frequency of occurrence of words in the particular event was proposed. The research on event extraction has strong domain relevance, emphasizing the extraction of relevant information from the text according to the specified event type and its template

In paper [8] the author has put forward a method for extracting events by taking frequently occurring named entities as core entities, which could perform cross-document event recognition tasks and then arrange the identified events on the timeline after it was performed the necessary calculations. The aim of the cross-document event ordering task is to build timelines from English news articles. To provide focus to the timeline creation, the task is presented as an ordering task in which events involving a particular target entity are to be ordered chronologically.

In paper [9] author proposed a method called TM-Gen, which is used for extracting information from any number of articles and representing them in a topic map format. Here they present an architecture to generate automatically a conceptual representation of knowledge stored in a set of text-based documents, they have used the topic maps standard and we have developed a method that combines text mining, statistics, linguistic tools, and semantics to obtain a graphical representation of the information contained therein, which can be coded using a knowledge representation language such as RDF or OWL. The procedure is language-independent, fully automatic, self-adjusting, and it does not need manual configuration by the user. Although the validation of a graphic knowledge representation system is very subjective, we have been able to take advantage of an intermediate product of the process to make an experimental validation of our proposal.

In paper [10] they described the process of using named entity recognition to obtain important definitions (citing different narrative styles of the same event) from news articles. Named Entity Recognition is a process where an algorithm takes a string of text (sentence or paragraph) as input and identifies relevant nouns (people, places, and organizations) that are mentioned in that string. News and publishing houses generate large amounts of online content on a daily basis and managing them correctly is very important to get the most use of each article.

Name Entity Recognition can automatically scan entire articles and reveal which are the major people, organizations, and places discussed in them. Knowing the relevant tags for each article help in automatically categorizing the articles in defined hierarchies and enable smooth content discovery.

In paper [11] the author describes on various different natural language processing techniques where these systems have been developed and utilized to extract events and clinical concepts form text, and several success stories in applying these tools have been reported widely. The most basic and useful technique in NLP is extracting the entities in the text. It highlights the fundamental concepts and references in the text. Named entity recognition (NER) identifies entities such as people, locations, organizations, dates, etc. from the text. Sentiment analysis is a part of NLP is most useful in cases such as customer surveys, reviews and social media comments where people express their opinions and feedback. The simplest output of sentiment analysis is a 3-point scale: positive/negative/neutral. In more complex cases the output can be a numeric score that can be bucketed into as many categories as required.

In paper [12] author proposed an approach that relies on both atrial and ventricular activity analysis, based on a novel non-linear filtering technique recently proposed to extract short-term events from biomedical signals. This is a fast novel non-linear filtering method named Relative-Energy (Rel-En), for robust short-term event extraction from biomedical signals. they developed an algorithm that extracts short- and long-term energies in a signal and provides a coefficient vector with which the signal is multiplied, heightening events of interest. This algorithm is thoroughly assessed on benchmark datasets in three different biomedical applications namely, ECG QRS-complex detection, EEG K-complex detection, and imaging photoplethysmography (iPPG) peak detection. Rel-En robustly extracted short-term events of interest. The proposed algorithm can be implemented by two filters and its parameters can be selected easily and intuitively. Furthermore, Rel-En algorithm can be used in other biomedical signal processing applications where a need of short-term event extraction is present.

There are techniques in NLP that help summarize large chunks of text. Text summarization is mainly used in cases such as news articles and research articles. Two broad

approaches to text summarization are extraction and abstraction. Extraction methods create a summary by extracting parts from the text. Abstraction methods create summary by generating fresh text that conveys the crux of the original text. There are various algorithms that can be used for text summarization like LexRank, TextRank, and Latent Semantic Analysis. To take the example of LexRank, this algorithm ranks the sentences using similarity between them. A sentence is ranked higher when it is similar to more sentences, and these sentences are in turn similar to other sentences.

Aspect mining of NLP identifies the different aspects in the text. When used in conjunction with sentiment analysis, it extracts complete information from the text. One of the easiest methods of aspect mining is using part-of-speech tagging. Aspect-Based Opinion Mining (ABOM) involves extracting aspects or features of an entity and figuring out opinions about those aspects. It's a method of text classification that has evolved from sentiment analysis and named entity extraction (NER). ABOM is thus a combination of aspect extraction and opinion mining. While opinions about entities are useful, opinions about aspects of those entities are more granular and insightful. The ABOM workflow constitutes initial text pre-processing, POS tagging, splitting sentences to extract aspects and classifying them into various dimensions/buckets.

According to WordNet [13], a general definition of a news event is “a specific thing happens at a specific place and time”, which may be consecutively reported by various media within a period. Most prevailing approaches to news event detection were proposed in this paper [14] they were mainly variants and improvements of the single pass method and agglomerative clustering algorithms.

The single pass clustering algorithms that scan the data only once can be classified into two categories: Partitional and Hierarchical. Partitional clustering attempts to directly decompose the data set in to a set of disjoint clusters. More specifically, they attempt to determine an integer number of partitions that optimize a certain criterion function. The criterion function may emphasize the local or the global structure of the data and its optimization process is an iterative procedure. Single pass K-means algorithm belongs to this category. Hierarchical clustering proceeds successively by either merging smaller cluster into larger ones, or by splitting larger clusters. Text clustering algorithms could be classified in several groups: vector space models, k-

means variations, generative algorithms, spectral algorithms, dimensionality reduction methods and phrase-based methods. Vector space model is a classic approach which shows better results on homogeneous topics and needs to know the number of clusters. K-means algorithm and its extensions are historically most popular approaches for hierarchical and partitioned clustering. However they have a number of drawbacks: effectiveness decreases on large data corpora and relies on random initialization. Also they are susceptible to outliers and noise and needs to know the number of clusters as well. Generative algorithms are also sensitive to outliers and it makes them less effective on heterogeneous data and have cluster count as input.

The basic idea of K Means clustering in text is to form K seeds first, and then group observations in K clusters on the basis of distance with each of K seeds. The observation will be included in the n^{th} seed/cluster if the distance between the observation and the n^{th} seed is minimum when compared to other seeds. *K-means* (KM) algorithm groups *N* data points into *k* clusters by minimizing the sum of squared distances between every point and its nearest cluster mean (centroid). This objective function is called *sum-of-squared errors* (SSE). Although k-means was originally designed for minimizing SSE of numerical data, it has also been applied for other objective functions sometimes the term *k-means* is used to refer to the clustering problem of minimizing SSE.

The vector space model [15] is an approach that reveals better results for homogeneous events and requires ensuring the number of clusters in advance. Vector space model is the most Widely used model used to represent document. In the statistically based vector-space model, a document is conceptually represented by a vector of keywords extracted from the document, with associated weights representing the importance of the keywords in the document and within the whole document collection; likewise, a query is modelled as a list of keywords with associated weights representing the importance of the keywords in the query. The weight of a term in a document vector can be determined in many ways.

In the paper [16] Some techniques related to ontologies, machine learning, and natural language processing were applied to help the documentalists of categorization and tagging of news, here we can get the desired news based on the keywords present and extract relevant information as the user needed. Many studies [17] have faced such problems, which have not yet been solved. Many researchers applied events detection to a specific field, working on detection

of economic events that may influence the market, such as mergers.

In paper [18] an event detection framework to discover real-world events from multiple data domains, including online news media and social media. Automatic keyword extraction is the process of identifying key paragraphs, key phrases, key terms, or keywords in an article that properly represent the document topic.

Due to keyword extraction provides a compact representation of article, some applications, such as automatic classification, automatic clustering, automatic indexing, automatic filtering, and automatic summarization, can benefit from the keyword extraction process [19]. Keyword is the smallest meaningful element of language that can move independently in Chinese, and the relationship between news and corresponding keywords has been studied extensively, falling into two categories: supervised and unsupervised approaches. In the former, the keyword extraction algorithm consists of two steps. The first step is training, in which models are trained to find keywords from labeled news reports. The second step is extraction, where the keywords are chosen from news reports that are not trained. The latter is composed of simple statistical approaches, linguistic approaches and machine learning concepts.

In paper [20] the author tried to extract keywords from micro blogs using a three-feature graph model, semantic space and word location. In paper [21] focused on a structure approach based on exploded events and graph generation. This paper proposed a method to find solutions for problems such as high variance and lexical variants. Kim [22] detected exploded and popular keywords including abbreviations. Zimniewicz [23] applied a scheduling model for a class of keyword extraction approaches and proposed methods for the overall performance evaluation of different algorithms, which are based on processing time and correctness (quality) of answers. Duari [24] proposed a parameter less keyword extraction method (sCAKE) based on semantic connectivity of words, combining with graph construction and scoring methods. In this paper, a multi-feature fusion will be used for keyword extraction.

CHAPTER 3

SYSTEM MODEL AND DEFINITION

Here in this chapter there's calculation of the similarity between news and events, there's transformation of news reports into a form that a computer can understand, that is, to build a model to represent news reports. The commonly used text representation model is the vector space model, which we also use in this paper. Every dimension of vector is a feature item extracted from news. Using vectors to represent text, the relevant knowledge in mathematics to calculate the similarity between vectors can be used, and the similarity between vectors can be referred to as the similarity between news, thus reducing the difficulty of calculating similarity between news reports.

Vector space model or term vector model is an algebraic model for representing text documents (and any objects, in general) as vectors of identifiers, such as, for example, index terms. It is used in information filtering, information retrieval, indexing and relevancy rankings. Its first use was in the SMART Information Retrieval System. Each dimension corresponds to a separate term. If a term occurs in the document, its value in the vector is non-zero. Several different ways of computing these values, also known as (term) weights, have been developed. One of the best known schemes is tf-idf weighting (see the example below). The definition of term depends on the application. Typically terms are single words, keywords, or longer phrases. If words are chosen to be the terms, the dimensionality of the vector is the number of words in the vocabulary (the number of distinct words occurring in the corpus). Vector operations can be used to compare documents with queries.

The vector space model has the following advantages over the Standard Boolean model:

1. Simple model based on linear algebra
2. Term weights not binary
3. Allows computing a continuous degree of similarity between queries and documents
4. Allows ranking documents according to their possible relevance
5. Allows partial matching

Most of these advantages are a consequence of the difference in the density of the document collection representation between Boolean and term frequency-inverse document frequency approaches. When using Boolean weights, any document lies in a vertex in a n-dimensional hypercube.

3.1 Comparison between news and event model

Here we collect all the news reports and event models, for each news report N, we can extract keywords from news headlines and contents as feature items and construct news vector models based on these characteristics. Using the vector $N(n_1,m_1,n_2,m_2,\ldots,n_k,m_k)$ to represent a news report, n is the feature of the vector, and the feature term is the keyword extracted from the news. The variable m_k is the weight of the feature item, these news reports collected here represents a collection of all the possible news happening, we then collect the possible event models also which is used to compare if the news reports are also represented as the hot events, for each hot event, a vector $E(e_1,s_1,e_2,s_2,\ldots,e_i,s_i)$ can be constructed to represent the event, of which e_i is a keyword extracted from the event and s is the weight of the keyword related to the event. We specify a ten-dimensional vector for an event; then, the initial elements of the vector are set to the same as the first news in this event. When subsequent news is classified to the event, keywords change, and the corresponding weights are recalculated.

This comparison mechanism we try to see if any news model is already classified as the hot event model, hot event models are considered the one which the end user is always interested to read. This type of procedure is really important as we get a featured extraction of news content which have the most priority news over the others. This procedure of comparison gives the result between 0 to 1, this is explained in detailed below with the related formula.

3.2 Similarity calculation

In order to support this functionality, which would be finding groups of reports describing the same event, we take the similarity comparison method into consideration. This section explains how to calculate an approximate similarity between news reports and events. Then, the computation is based on an aligned set of basis vectors obtained using keyword correlation

analysis. In this paper, we propose a new method to calculate the similarity between news and events. The following is the relevant definitions for the calculations :

Definition N: N represents a piece of news.

Definition E: E represents an event.

Definition t_0 : t represents the current time.

Definition P_1: P is a set, $P = \{k1,k2,\ldots,ki\}$, one of which k represents a keyword that appears at the same time in news N and event E.

Definition P_2: P is a set, $P = \{w1,w2,\ldots,wi\}$, which contains the weight for each of the keywords

Definition P_3: P is a set, $P = \{t1,t2,\ldots,tp\}$, which contains the last time that each keyword in P was recently updated.

Definition P_4: P is a set, $P = \{e1,e2,\ldots,ep\}$, which contains all keywords for an event.

Definition P_5: P is a set, $P = \{s1,s2,\ldots,sp\}$, which contains the weight of each keyword in P.
Definition P_6: P is a set, $P = \{q1,q2,\ldots,qp\}$, which contains the last time that each keyword in P was most recently updated.

The similarity calculation formula is :

$$Sim(N,E) = \frac{\sum_{i=1}^{m} \frac{1}{t_0 - t_i} \times w_i}{\sum_{j=1}^{n} \frac{1}{t_0 - q_j} \times s_j}$$

The similarity value obtained with the above formula is a fractional number between 0 and 1. Depending on the threshold we decide whether the news report and the event model are same, if the value obtained is near to 1 then there's a greater possibility that the news N and event E is similar or if any missing facts in event E then the news N can be classified into event E. In the converse way, if the similarity value obtained in the above formula is near to 0 then the news N and event E are different and news can't resemble the event. Basically, there's a certain threshold which is used for these calculations. If the value is larger than a certain threshold, then the news is classified into the event. Otherwise the news is stored as a new event in the event library.

This similarity index plays an important role in comparison between the news model and event model, we get to better classify these events. If multiple news articles resemble the a certain event we can reduce the duplicate copies of these news articles and classify as a single event model. Event model plays an major role as they're highlighted for the users to see based on their interest.

3.2 Proposed MFE Scheme

Here we discuss on the concept of multi-feature keyword extraction scheme. We use two main concepts here that is the frequency and part of speech. TF is an acronym for term frequency, it describes the number of times and a particular term occurs in an article. It is generally believed that the higher the term frequency of a word, the more important the word is in the article. In the search engine optimization, a larger number of duplicate words are packed in one article. In order to escape this type of scenario we have set C_{total} which has total number of words in the news set, and $L_i = TF / C_{total}$. When $L_i > L$, we regard it as this word is completely useless Information with low importance, here L is set to be 0.75 .

Depending on the characteristics we have classified whether the words is usefull or not, this is shown below:

$$TF_{new} = \begin{cases} TF & \frac{TF}{C_{total}} \leq L \\ 0 & TF/C_{total} > L \end{cases}$$

According to this classification, if TF / C_{Total} is leff than L then the word is complete important and it's not classified as a duplicate word if the ratio is greater than L then the word doesn't have any importance.

According to the characteristics of the Chinese language, the keywords are generally Nouns (n) and Verbs (v), and a few adjectives and adverbs are included. Prepositions and auxiliary words generally cannot express specific meaning. The Names (nr), Place Names (nt) and Institution Names (ns) are more likely to become keywords. Therefore, this paper uses part of speech (POS) to adjust the weights of keywords. These keywords make an important role for the MFE algorithm. we actually group these names into set which is a vector.

Set A is a vector $\{nr,nt,ns,v\}$, t is the keyword, P_{weight} is the part of speech weight of words, p is the part of speech of candidate keywords, T is the keyword set, and P comprises the weight of the part of speech corresponding to the keywords. Adjustable variables a, b and c have general values 3, 2 and 1, respectively.

When $\forall p \in A \; and \; p = nr$,

$P_{weight} = TF_{new} * a$;

when $p = (ns|nt)$,

$P_{weight} = TF_{new} * b$ or when $p \in n \cap p \notin A \; or \; p = v$,

$P_{weight} = TF_{new} * c$, we add t to set T, and P to set P. The final set T is the filtered keyword set, and P is the part of speech weight of corresponding words.

Taking the consideration of a sample news report, keyword extraction is performed and the result will be the combination feature TF_{new} and POS (Part of Speech) is significantly better than the single TF_{new} feature. Under the influence of multi-feature keyword extraction, many irrelevant words were successfully removed.

TFnew and part of speech are two important features in evaluating the importance of keywords. According to the different degrees of importance of the two characteristics in the evaluation of keywords, we give the following formula :

$$W_i = k_1 TF_{new} + K_2 P_{weight}$$

In the above formula k_i is an adjustable parameter, generally 0, 1, 2 or 3. In the process of keyword extraction, TF features and part of speech features were integrated into the unified multi-feature fusion model to determine the weights of the keywords. Then the multiplying parameters a/b/c are used to emphasize the possibility of different words being the keyword according to part of speech, and the integration of Pweight and the TFnew is used to balance the relative importance between the two features of TFnew and POS, then keyword extraction for different types of text is achieved by adjustment of a/b/c and k_i .

CHAPTER 4

AI DRIVEN MFE SCHEME ANALYSIS

Here in this chapter we try the artificial intelligence driven Big Data MFE (Multi-feature extraction scheme) is constructed. We normally have a various source of unstructured information like from various channels such as the internet, social media and mobile networks, this information is very critical. The algorithms for the MFE scheme are explained in detailed below:

Algorithm 1 An AI-Driven Big Data MFE scheme

Input: a set $D\{d_1, d_2, \ldots, d_n\}$ of texts
Output: a set $F\{f_1, f_2, \ldots, f_m\}$ of features; a set $G\{g_1, g_2, \ldots, g_m\}$ of evaluation measures

1: **For i = 1: n do**
2: Extract keywords form d_i using MFE
3: **End for**
4: **For j= 1: m do**
5: Compute the feature data f_j
6: Cluster the texts D with single pass algorithm in a particular order associated with feature f_j
7: Compute evaluation measures $g_j\{v_1, v_2, v_3, v_4\}$
8: **End for**
9: **return** $G\{g_1, g_2, \ldots, g_m\}$

The above algorithm takes the input of texts in the form of set D and the output would be the features obtained from the set. The end condition of the algorithm is that no new features appear. In the specific application of MFE scheme, it can be summarized as the following fourn general procedural steps.

Step 1: Data Collection

In this data collection phase the primary focus was to gather information in support of our information security risk assessment. Without adequate data, there is very little value to the risk

assessment. The content of data collection covers all aspects of human activities such as scientific research, life, work and entertainment, and its forms include news, blog, BBS and microblog.

Step 2 : Data cleaning

Data cleaning is the process of ensuring that your data is correct, consistent and useable, It removes major errors and inconsistencies that are inevitable when multiple sources of data are getting pulled into one dataset. Using tools to cleanup data will make everyone more efficient since they'll be able to quickly get what they need from the data. Fewer errors means happier customers and fewer frustrated employees. The ability to map the different functions and what your data is intended to do and where it is coming from your data.

In the HTML raw text crawled by the crawler, data cleaning is required to filter out the label text. There are a lot of unnecessary information in the web page, such as advertisements, navigation bars, html code, javascript code, comments, etc. Information that we are not interested in can be deleted. If the main body extraction is required, the text can be extracted by using tag usage, tag density determination, data mining idea, visual web page block analysis technology and other strategies. Text data (usually spoken text records) may contain human emoji, such as [laughing], [Crying], [Audience paused].

These expressions are usually unrelated to what is being said and therefore need to be removed. This can be done with simple regular expressions. In addition, text data that people generate on social forums is essentially informal. Most tweets come with lots of sticky words like "RainyDay", "PlaingInTheCold" and so on. These also can be split into normal forms with simple rules and regular expressions, and all punctuation should be treated as a priority. For example, periods, commas, question marks are important punctuation that should be retained, while others need to be removed.

Step 3: Data Preparation

Data preparation can be divided into three parts: word segmentation, part-of-speech tagging, and text vector calculation. The data interaction between each step process is conducted through some data record files, so as to avoid the memory overflow error caused by adding all process data into memory at one time, for Chinese text data, such as a Chinese sentence, the words are continuous,

and the minimum unit granularity of data analysis we want is words, so we need to work on word segmentation, so that we can prepare for the next step. The purpose of part-of-speech tagging is to allow the sentence to incorporate more useful language information into subsequent processing. Text vector calculation is prepared for the subsequent calculation of word weight and the screening of keywords.

Step 4: Algorithm usage

In a single feature, the algorithm arranges texts in descending order according to the feature values, and prioritizes text with significant features. Texts with a large amount of reading, more comments, or faster commentary speeds can better represent the hot spots of public concern, that is, where social hot spots are. After all the features are traversed, the clustering results are evaluated according to the common evaluation indicators of text analysis, such as recall and precision, and the optimal results and their corresponding features can be obtained**.**

CHAPTER 5

HOT EVENT DETECTION

In this chapter there discussion on hot event detect algorithm, we use the keyword notation as the basis for this algorithm. With the help of the calculation of the similarity between the news reports and existing events, we can try to obtain the maximum value. If the max is larger than a given Similarity Threshold (ST), we assume that this news could be classified into the corresponding event. However, if it is not, we would create a new event based on this news in the range of the Proportion of Prepositive Clustering (POPC) or abandon it out of range. Afterwards, the weights of the keywords on the event must be handled carefully. Then, in descending order, according to the number of comments attached to the news, the algorithm could handle massive online news effectively

5.1 Hot Event Detection algorithm

The algorithm is shown below:

Algorithm 2 Hot Event Detection

Input: a set $S\{s_1, s_2, \dots, s_n\}$ of news reports
Output: a set $E\{e_1, e_2, \dots, e_m\}$ of events

```
For each news report s_i in S by a particular order
  If i == 1 then
    Set e_1 = s_1
    Add e_1 to E
  Else if i < POPC * size(S) then
    If sim(s_i, e_k) > ST then
      Add s_i to e_k
      Recalculate the weight of e_k
    Else
      Create a new event e_c
      Add e_c to E
    End if
  Else if sim(s_i, e_k) > ST then
    Add s_i to e_k
    Recalculate the weight of e_k
  Else
    Abandon the news report s_i
  End if
End for
Return E
```

There are seven steps for the complete algorithm.

Step 1: collecting

Here we try to collect all the possible news reports by crawling the news portals with a Web crawler. A web crawler (also known as a web spider or web robot) is a program or automated script which browses the World Wide Web in a methodical, in an automated manner to find the news articles. This process is called Web crawling or spidering. Many legitimate news sites, in particular news finding engines, use spidering as a means of providing up-to-date latest news data. Web crawlers are sometimes used to create a copy of all the visited pages for later processing by a news search engine, that will index the downloaded pages to provide fast searches. crawlers can be used to gather specific types of information from Web pages, where in this case, we try to find the news. In addition, extract keywords from every piece of news and simultaneously set the Number of Keywords (NOK). Then, news will be represented as a fixed length list of keywords from the content of the news itself.

Step 2: Sorting

Sort all news in descending order according to the number of comments attached to the news. We actually gather all the news reports and find its comments, posts views and taking all these parameters we try to order the news reports in the descending order.

Step 3: Initializing

Here we try to set the first piece of news with the most comments as the initial event and the news weights as that of the event, for the first initial step.

Step 4: Fetching

Here we fetch a news report from the news corpus in order and calculate the maximum similarity between the news and all the events based on the list of keywords. In this work, we compare the news obtained with the first event and obtain a similarity value. However, we are not sure whether it is the maximum or not, and we then calculate another similarity value between the news and the next event so that we can record the larger one and its corresponding event. With that

methodology, the maximum similarity can be calculated. This step carries out in every iteration to compare the similarity between the news reports and the events model.

Step 5: comparison

If the maximum similarity between the news report and events model, is larger than the given similarity threshold, we classify this news to the corresponding event and add the weights of same keywords representing news to the corresponding weights of the event. Then, divide all the keyword weights of the event by the number of news items belonging to this event, and the newly calculated event keywords are obtained. After that, if the weight of the news keywords is larger than that of the event, the algorithm will replace the smaller weight and its keyword with the larger pair. Besides, we assume that news with a larger number of comments is most likely to be an event. And In the initial cluster, the news with the most comments firstly constitute the initial event set, so for subsequent news, which keywords haven't appeared in the key words of events, the algorithm will directly remove it.

Step 6: condition

If the maximum similarity between the news report and event model, is not larger than the given similarity threshold, a new event for the news is created, and the keywords and weights of the news are regarded as the event's, with the news report in the range of POPC. In addition, the news report will be abandoned if it is out of range of that value.

Step 7: repetition

Repeat the processing steps from step 4 to step 6 until all the news is handled.

There are three impact factors in our algorithm that may influence the result of news clustering, including the number of keywords, the similarity threshold and the proportion of prepositive clustering. One of the key issues in this algorithm is the similarity calculation, which is based on keywords representing news. According to the formula calculating similarity, we assume that the increment or decrement to the number of keywords directly changes the result of the similarity calculation, thereby possibly impacting the maximum similarity.

In order to handle this, simple comparison of the max to the similarity threshold is performed to determine whether the news is classified to an existing event stored in the event library or not. It will be hard to classify the news to a certain event if the similarity threshold is too high. Thus, the similarity threshold will have a significant impact on the result of news clustering.

Through the initial local clustering, the algorithm first automatically generates a hot event library in which the global number of hot events and their contents are determined by the proportion of prepositive clustering. If POPC has a higher proportion, more news will be clustered with more events extracted. In this paper, the proportion of prepositive clustering is also explored.

CHAPTER 6

EXPERIMENT ANALYSIS

In this chapter, there's evaluation of our proposed approach to learn the event mining. We use certain tools like excel to plot figures and python platform to implement all algorithms. The compared approaches are respectively classical clustering algorithms and the multifeature model.

6.1 Data Preparation

Our huge set of data is built from the NetEase news portal, which contains 19681 news articles from July 1 through August 9 of 2017. Each piece of news is marked as on-event or off-event for every one of the events extracted artificially. We defined a piece of news as on-event if it is related to one of the standard events and off-event if it is not. Taken absolutely, an on-event news article belongs only to a certain event, not to two or more. The below table lists some statistics about all the standard events and shows the number of news belonging to each event.

Table – 6.1 Events Description

Events	The Number of News
Court to Freeze Assets of LeEco Founder Jia Yueting.	504
Baoding Rongda threatens to quit Chinese league after controversial draw.	110
Logistics War between Jingdong and Suning.	37
Chinese teacher from Fujian traveling in Japan has gone missing.	32
Zou Shiming loses the WBO flyweight title as he's stunned in the 11th round by Japan's Sho Kimura.	54
Earthquake strikes China's remote Sichuan province.	258
Death of university graduate sparks anger at Chinese pyramid scam gangs.	129
Actor Xu Zheng was exposed and wounded female.	10
The resignation of Anti-GMO activist Cui Yongyuan.	8
Wolf Warrior 2 is the highest grossing movie in the world, beating out Hollywood blockbusters and epic European sci-fi Valerian.	103

6.2 Evaluation Measures

We have certain evaluation measures considering it's relevance was used as the main option, the evaluation indices system for testing the efficiency of clustering news were established, and it's four components were the generation rate, precision, recall and F1 – score.

In the table 6.1, the sum of all the events in the standard event set is 10, which is marked as a constant variable n. we try to let the set of n standard events E_n be { E_1, E_2,, $E_{3....}$ E_n } and the set of events detected automatically be T_r = {T_1,T_2, $T_{3.....}T_r$}, of which each T consists of keywords {$T_{1.....}T_k$} as well as each E = { $e_{1....}e_k$ } and k is below limit ten. If there is an event E_i that has a percentage of the same keywords between E_i and T_j greater than a 30%, we define the generation rate could be expressed as follows:

$$generation\text{-}rate = \frac{|E_n \cap T_r|}{n}$$

where it reaches its best value at 1 and its worst value at 0. In this paper, we define precision and recall based on the standard events and clustering events detected by the method we propose. Precision (also called positive predictive value) is the fraction of relevant instances among the retrieved instances. Precision is used with recall, the percent of *all* relevant news documents that is returned by the search. The two measures are sometimes used together in the F1 Score (or f-measure) to provide a single measurement for a system. Precision takes all retrieved documents into account, but it can also be evaluated at a given cut-off rank, considering only the topmost results returned by the system. This measure is called *precision at n*, the formula for precision is given below:

$$precision = \frac{\sum_{i=1}^{n} \frac{|E_i \cap T_k|}{|E_i|}}{n}$$

Recall (also known as sensitivity) is the fraction of the total amount of relevant instances that were actually retrieved. Both precision and recall are therefore based on an understanding and measure of relevance.

$$recall = \frac{\sum_{i=1}^{n} \frac{|E_i \cap T_k|}{|T_k|}}{n}$$

where if an event T exists in the set of events Tr and it makes $E_i \cap (\forall T \in Tr)$ take the maximum value, we regard it as Tk. Intuitively, precision is the ratio of events that are clustered.

To combine the two indicators mentioned above, we take the F1-score into account. In the equation below, the precision and recall are weighted equally

$$F_1\text{-}score = \frac{2*precision*recall}{prcision+recall}$$

Thus, the closer the F1-score is to 1, the higher the clustering quality.

6.3 Experimental Design

There are two sets of experiments were performed in our proposed work. The first experiment investigated the measures for evaluating our approach with the aforementioned dataset. In this experiment, there were a total of 224 parallel tests, and among them, differentiators depended on three variable parameters, including the number of keywords, similarity threshold and proportion of prepositive clustering.

We set the number of keywords representing a news article from 4 to 18 with the specific interval 2, the similarity threshold is ranging from from 0.1 to 0.7 with the specific interval 0.1, and the proportions of prepositive clustering at 1%, 3% , 5% and 7%, yielding the 224(8*7*4) pairs of different combinations, For an example, in one test, we clustered the news based on the number of keywords as 6, a similarity threshold of 0.3 and the proportion of prepositive clustering at 1%. In the second experiment, we examined the performance of the classical clustering algorithms on the same set of data.

In order to compare our method, we try to choose few clustering algorithms as the baseline. In k – means algorithm, news articles are places into k partitions, and the initialization methods Forgy and Random [22] partitions are used. the Forgy method randomly chooses k observations

from the corpus and uses these as the initial means, while the Random Partition method first randomly assigns a cluster to each observation and proceeds to the update step and computing the initial mean to be the centroid of the cluster's randomly assigned points until it is improved. We chose initial centers in a way that gives a provable upper bound on the WCSS object, and the filtering algorithm used kd-trees to speed up each k-means step. Besides, mini batch k-means, dbscan clustering, birch clustering, spectral clustering, agglomerative clustering, mean shift clustering and affinity propagation clustering are also on the list.

6.4 Experimental Results and Analysis

As this can be seen from the below figures, the linear correlations of the data are similar in shape. The data using ten keyword representations appear balanced and a little superior to the data using other keyword representations. This phenomenon reflects the small amount of available information, which means that using ten keyword representations could lead to better clustering results.

The below graph shows different representations of news in terms of generation rates, precision, recall and F1 – score.

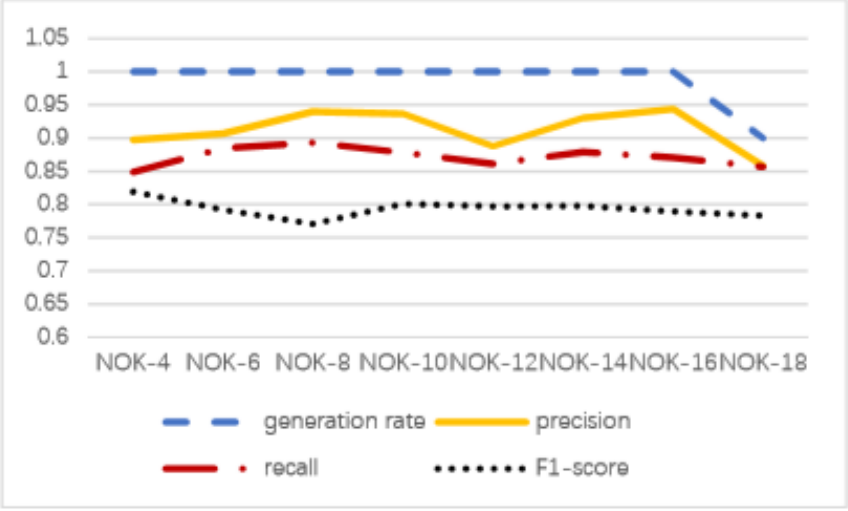


Figure 6.4.1 Representations of news in terms of generation rates, precision, recall and F1 – score

The below figures from 6.4.2 to 6.4.6 show the effectiveness of the approach proposed in this work with various parameters, comparing the generation rate, precision, recall and F1 -score. These below figures signify the line extending outwards, the better the clustering results. The graphs here have two preliminary conclusions.

The first conclusion is for the proportion of prepositive clustering, they have found that the highest percentage of 7% outperforms any other one. Furthermore, the combination of POPC 7% and ST 40% had greater influence on the experimental results.

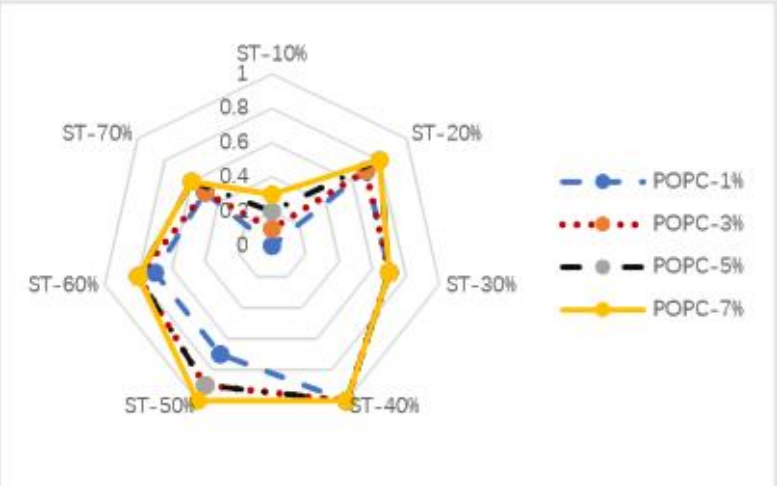


Figure 6.4.2 Tradeoff between POPC and ST in terms of generation rates

In the figure 6.4.2 this shows POPC and ST in terms of generation rates where number of keywords is ten.

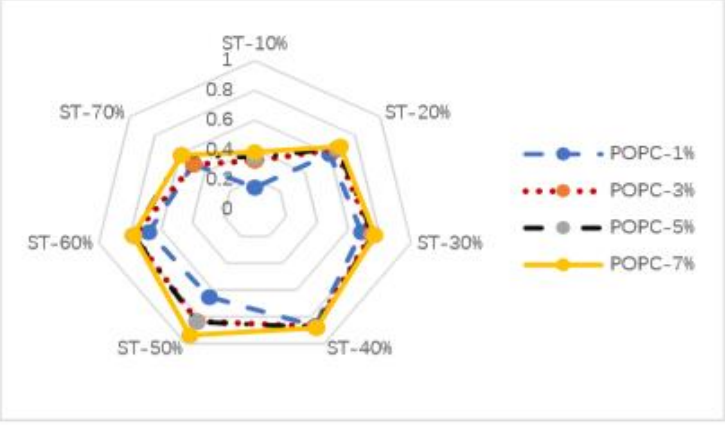


Figure 6.4.3 Tradeoff between POPC and ST in terms of precision where NOK is ten

In the above figure we can see the tradeoff between POPC and ST in terms of precision values, where the number of keywords is ten.

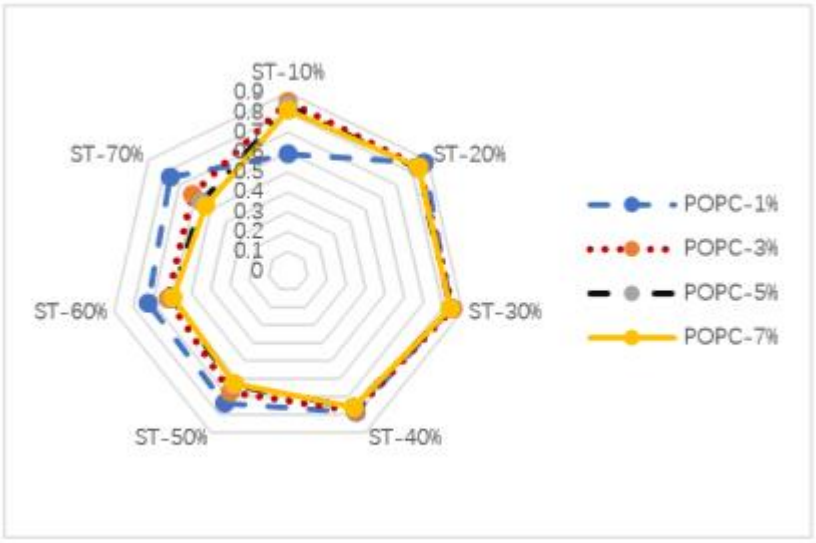


Figure 6.4.4 Tradeoff between POPC and ST in terms of recall, where NOK is ten

In the above figure 6.4.4 we try to compare the values between in POPC and ST in terms of recall, where NOK is ten.

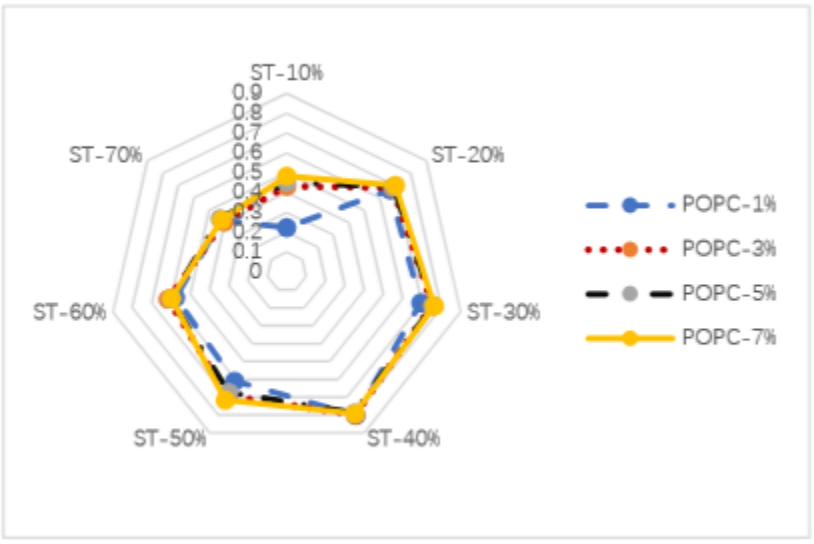


Figure 6.4.5 Tradeoff between POPC and ST in terms of F1 -score, where NOK is ten

A certain set of event keywords extracted by the hot event detection algorithm are listed in the

below table and the listing order corresponds to the standard events.

There could be any type of keywords extracted like sports, countries keywords, names, institute names, places, various important things. Some of the keywords of certain events are shown below for POPC is 7% and ST is 40%.

Table 6.2 Events keywords

Events No.	Event keywords
1	Letv, Jia Yueting, corporation, freeze, holding, justice, supplier, media, financial institution
2	club, football, soccer fans, competition, event, Baoding, quit, punishment, hope
3	expressage, China, physical distribution, Jingdong, Suning, e-commerce, Tonglu County, service, speed
4	Wei Qiujie, Japan, loss of communication, discover, journey, journalist, hotel, Fujian, teacher
5	Zou Shiming, Kimura, opponent, bout, boxing, offensive, world, occupation, stamina
6	earthquake, photograph, net friend, China, happen, aba prefecture, Jiuzhaigou county, Sichuan Province
7	pyramid scheme, activity, crime, Ministry of Public Security, public security organ, Shan Xin Hui, pursuant to the law, mastermind, organization, safeguard
8	female, reconciliation, Xu Zheng, deny, expose, exclusive, video, injury, actor
9	Cui Yongyuan, store, food, transgenosis, position, interest group, offend, statement, resign
10	passport, Wolf Warrior, movie, Wang Cailiang, box office, China, market, corporation, theme

In table 6.2 we gathered keywords from 10 events which contains the top keywords that identify the events. These gather the most important part of the article.

In the table 6.3 we gather the experimental results for POPC 7% and NOK 10%.

Table 6.3 Experimental results for POPC 7% and NOK 10%

	precision	recall
ST-10%	0.385612102	0.812311218
ST-20%	0.680086392	0.840866789
ST-30%	0.768732519	0.837736526
ST-40%	0.883824456	0.759592619
ST-50%	0.936733762	0.626990142

Here in the table 6.3 we have collected the values of precision and recall with similarity threshold for every 10% , the precision seems to be highest at similarity threshold at 50% and it differs for recall.

The second conclusion is having the direction of the similarity threshold, an inverse relationship between he precision and recall is evidenced, where it is possible to increase one at the cost of reducing the other. the similarity threshold.. Usually, precision and recall are not dissected and discussed in isolation. Instead, the measure that is a combination of precision and recall is the F1-score (the weighted harmonic mean of precision and recall), which is more useful for reference and often used in the field of information retrieval for query classification performance. We will further explore the approach to verify whether the performance improved from this approach by setting up other experiments for comparison. In this work, we regard the keyword vector extracted by the general feature of TF-IDF and MFE schema as the inputs for clustering algorithms baselines. The comparison results between the clustering algorithms and the multi-feature model are shown in below figure 6.4.6.

Unfortunately, there are five algorithms (including birch clustering, spectral clustering, agglomerative clustering, mean shift clustering and affinity propagation clustering) failing to cluster with same hardware, and the reason for the failure is there is no enough memory. It means these algorithms are memory intensive compared to our approach.

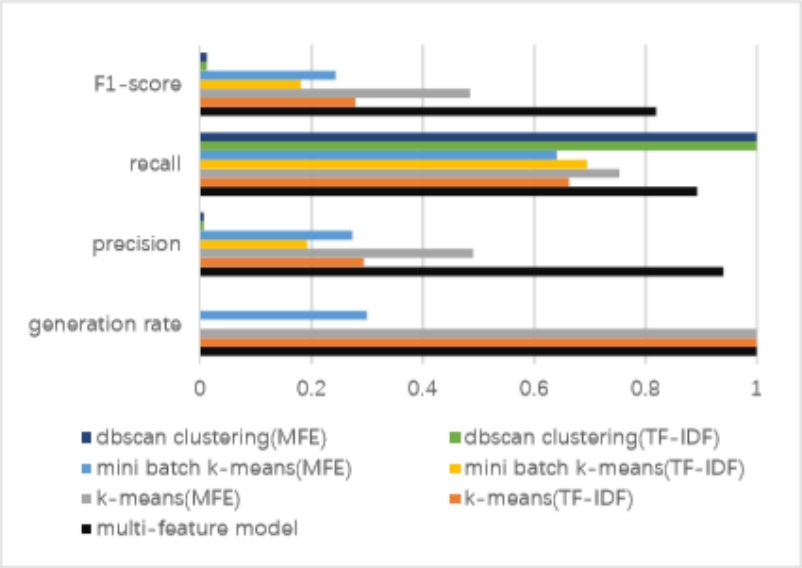


Figure 6.4.6 experimental results of various approaches

There are significant differences in the above figure 6.4.6, seven groups of experiments, and our approach achieves the optimized result as well as the black part. The results demonstrate that our approach offers good performance and MFE schema is superior to TF-IDF as input.

CONCLUSION AND FUTURE ENHANCEMENT

In this Topic, we proposed a multi-feature keyword extraction method, based on which we designed an artificial intelligence driven big data MFE scheme and expanded an application example of this universal scheme. We have discussed the task of clustering algorithms with application to news overloading on the Internet. Through the experiments we designed, we conducted a detailed empirical study on the feasibility and validity of our approach. According to the analysis of the experimental results, we can obtain a better event generation rate, precision, recall and F1-score when using the specified POPC, ST and NOK, which provides an effective clustering method for detecting hot events from the massive amount of online news. Thus, for practical applications, the approach provides a reference value. However, this method also has the following shortcomings. The threshold set in this paper is a fixed value, while the news flow is dynamic data. If the threshold can be dynamically adjusted according to the event detection, the detection effect of the hot event can be improved. Moreover, the source of hot events is limited to news only. However, in daily life, many hot events may first appear in microblogging, forums and so on. If we want to provide users with more comprehensive social hot events, the source of information should not be confined to the news, but should be integrated with multiple data sources. We will solve these problems in future work and study the feasibility of applying this method to knowledge discovery.

CHAPTER 7

REFERENCES

[1] Si, H., Chen, Z., Zhang, W., Wan, J., Zhang, J., & Xiong, N. N. (2019). A member recognition approach for specific organizations based on relationships among users in social networking Twitter. Future Generation Computer Systems, 92, 1009-1020.

[2] Nie, Z., Wen, J. R., & Yang, L. (2012). U.S. Patent No. 8,229,960. Washington, DC: U.S. Patent and Trademark Office.

[3] Alani, H., Kim, S., Millard, D. E., Weal, M. J., Hall, W., Lewis, P. H., & Shadbolt, N. R. (2003). Automatic ontology-based knowledge extraction from web documents. IEEE Intelligent Systems, 18(1), 1421.

[4] Suchanek, F. M., Kasneci, G., & Weikum, G. (2007, May). Yago: a core of semantic knowledge. In Proceedings of the 16th international conference on World Wide Web (pp. 697-706). ACM

[5] Gkatziaki, V., Papadopoulos, S., Mills, R., Diplaris, S., Tsampoulatidis, I., & Kompatsiaris, I. (2018). easIE: Easy-to-use information extraction for constructing CSR databases from the web. ACM Transactions on Internet Technology (TOIT), 18(4), 45.

[6] Kluegl, P., Toepfer, M., Beck, P. D., Fette, G., & Puppe, F. (2016). UIMA Ruta: Rapid development of rule-based information extraction applications. Natural Language Engineering, 22(1), 1-40.

[7] Doddington, G. R., Mitchell, A., Przybocki, M. A., Ramshaw, L. A., Strassel, S. M., & Weischedel, R. M. (2004, May). The Automatic Content Extraction (ACE) Program-Tasks, Data, and Evaluation. In LREC (Vol. 2, p. 1).

[8] Luhn, H. P. (1958). Auto-encoding of documents for information retrieval systems. IBM

Research Center.

[9] Garrido, A. L., Buey, M. G., Escudero, S., Ilarri, S., Mena, E., & Silveira, S. B. (2013, November). TM-gen: from text documents. In 2013 IEEE.

[10] Shinyama, Y., Sekine, S., & Sudo, K. (2002, March). Automatic paraphrase acquisition from news articles. In Proceedings of the second international conference on Human Language Technology Research (pp. 313-318). Morgan Kaufmann Publishers Inc.

[11] Manning, C. D., Manning, C. D., & Schütze, H. (1999). Foundations of statistical natural language processing. MIT press.

[12] Yazdani, S., Fallet, S., & Vesin, J. M. (2018). A novel short-term event extraction algorithm for biomedical signals. IEEE Transactions on Biomedical Engineering, 65(4), 754-762.

[13] Dave, K., Lawrence, S., & Pennock, D. M. (2003, May). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In Proceedings of the 12th international conference on World Wide Web (pp. 519-528). ACM.

[14] Brants, T., Chen, F., & Farahat, A. (2003, July). A system for new event detection. In Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval (pp. 330-337). ACM.

[15] George, A. (2013). Efficient high dimension data clustering using constraint-partitioning k-means algorithm. Int. Arab J. Inf. Technol., 10(5), 467-476.

[16] Salton, G., Wong, A., & Yang, C. S. (1975). A vector space model for automatic indexing. Communications of the ACM, 18(11), 613-620.

[17] Lefever, E., & Hoste, V. (2016). A classification-based approach to economic event detection in dutch news text. In Tenth International Conference on Language Resources and Evaluation (LREC'16) (pp. 330-335). European Language Resources Association (ELRA).

[18] Yang, Z., Li, Q., Wenyin, L., & Lv, J. (2019). Shared Multi-view Data Representation for Multi-domain Event Detection. IEEE transactions on pattern analysis and machine intelligence.

VISVESVARAYA TECHNOLOGICAL UNIVERSITY

"JnanaSangama", Belagavi-590 014, Karnataka, India.

TECHNICAL SEMINAR REPORT

ON

IRIS RECOGNITION AS A BIOMETRIC TECHNIQUE

Submitted in partial fulfilment for the award of the degree of

BACHELOR OF ENGINEERING
In
Electrical & Electronics Engineering

Submitted by

RAMAKRISHNA A (1AM17EE402)

Under the guidance of

Mrs R SELVAMATHI
Associate Professor, EEE Dept.
AMCEC, Bangalore

AMC Engineering College

Department of Electrical and Electronics Engineering

Bengaluru- 560 083.

2019-20

AMC ENGINEERING COLLEGE

18 KM, BANNERGHATTA ROAD, BANGALORE-560083

DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING

CERTIFICATE

This is to certify that the technical seminar report entitled "**IRIS RECOGNITION AS A BIOMETRIC TECHNIQUE**" Carried out by **RAMAKRISHNA A** a bonafide student of AMC Engineering College, Bangalore, in partial fulfilment for the award of the degree of the Bachelor of Engineering in Electrical and Electronics Engineering, of the **Visvesvaraya Technological University, Belagavi** during the year **2019-20.** It is certified that all the correction/suggestions indicated for internal assessment have been incorporated in the report deposited in the department library. The internship report has been approved as it satisfies the academic requirements with respect to the internship work of prescribed for the said degree.

Signature of Internal guide	**Signature of HOD**	**Signature of the Principal**
(Prof. R Selvamathi)	(Dr. K.N Bhanuprakash)	(Dr. A.G Nataraj)

DECLARATION

I, RAMAKRISHNA A, student of Electrical and Electronics Engineering, AMC Engineering college Bengaluru, hereby declare that the internship work entitled **"IRIS RECOGNITION AS A BIOMETRIC TECHNIQUE"** has been carried out at AMC Engineering College, under the guidance of **Prof**. **R SELVAMATH**I, Associate Professor, Dept. Of Electrical and Electronics Engineering, AMC Engineering College, Bengaluru and submitted in partial fulfilment of the course requirements for the award of the degree in **Bachelor Engineering in Electrical and Electronics Engineering from the Visvesvaraya Technological University, Belagavi,** during the year **2019-20**.

We also declare that, to the best of our knowledge, work reported here is not a part of any other dissertation on the basis of which a degree or award was conferred on an earlier occasion on this, by any other student.

Date:

Place: Bengaluru **RAMAKRISHNA A**

ACKNOWLEDGEMENT

I would like to thank our chairman **Dr.K .R. Paramahamsa** and CEO, **Dr. T. N. Sreenivasa**, AMC Engineering College, Bangalore, for providing the necessary infrastructure.

I would like to thank our Principal, **Dr. A.G. Nataraj**, AMC Engineering College, Bangalore, for his kind cooperation.

I sincerely thank **Dr.Bhanuprakash.K.N**, HOD, Dept. of Electrical and Electronics Engineering, AMC Engineering College, Bangalore, from the bottom of my heart for his support and understanding.

I consider it my privilege to express gratitude to my guide, **Prof**. **R SELVAMATHI**, Associate Professor, Dept. Of Electrical and Electronics Engineering, AMC Engineering College, Bangalore, without his support and guidance, this internship report would not have been a success.

My deepest gratitude goes to my teachers, who showed infinite patience and understanding till the completion of my technical seminar.

Last, but not the least, my sincere credit to my parents, my friends and to one and all who have directly or indirectly helped me in the successful completion of the technical seminar.

RAMAKRISHNA A(1AM17EE402)

ABSTRACT

In this technical seminar, we have understood and learnt about Several safety regulations particularly in the charging electric vehicles (EVs) are developed to ensure the electric safety and prevent the hazardous accidents, in which safety requirements for electric vehicle supply equipment (EVSE) and the EV battery.Quantitative assessment of electrical safety considering the operation conditions of large-scale electric vehicle charging stations (EVCSs).Evaluate the electrical safety of the large scale EVCSs when coupled to renewable power generation. Fundamental electrical safety issues, and the protection against electric shock of persons interacting with electrical vehicles.

CHAPTER-1

INTRODUCTION

1.1. BIOMETRIC TECHNOLOGY

A biometric framework gives automatic recognition of an individual based on certain unique characteristics or feature possessed by them. Biometric frameworks have been developed based on fingerprints, facial elements, voice, hand geometry, handwriting, the retina, and the iris.

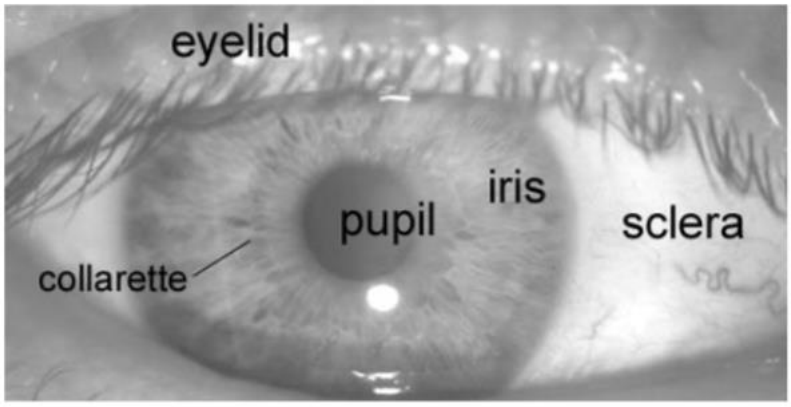
The biometric framework works by: Capturing a specimen of unique feature Transforming the specimen using couple of numerical models into biometric layout This biometric format will provide a standardized, efficient and profoundly segregating portrayal of feature d. Comparison with other layouts to determine identity

A decent biometric is described by utilization of an element that is; thoroughly unique – so that the possibility of any two

individual having a similar characteristic will be insignificant, immutable – so that the feature remains unfluctuating over the period of time, and be adequately obtained– so as to provide suitability to the user, and avert dispersion of the feature.

1.2. The Human Iris

Iris is the pigmented region of the eye. It is a circular sinewy diaphragm separating the two regions of the eye. It extends from ciliary muscle across the eyeball in front of the lens. It has a small circular aperture in the middle through which the light enters the eye, which is called pupil. The iris controls the amount of light entering the eye by contracting or relaxing the eye muscle, and hence contracting or dilating the pupil.



[Figure 1] Iris

The particular pattern in the iris region is formed during the elementary term of life, and stromal pigmentation occurs in the following couple of years.The incidental process of formation of the unique patterns of the iris is not related to any genetic factors. The only characteristic that depends on ancestral genes is the pigmentation of the iris, giving eye its color. As a result leading to an autonomously independent pattern of the two eyes of an individual. Furthermore, identical twins acquire non-germane iris patterns

1.3. Iris Detection

The iris is a well-protected part of the eye, although it is externally visible whose unique self-generated pattern remains stable throughout adult life. These key factors which make the Iris suitable as a biometric for identifying individuals.

Image processing frameworks can be used unique feature and pattern extraction along with converting it into the biometric template from the digital image of the eye, which can be later stored in the database. This biometric template contains a physical-mathematical representation of the unique information stored in the iris and allows comparisons to be made between models.

When a client prefers to be distinguished and identified by an iris recognition system:

- The image of eye needs to be acquired and is photographed(Image acquisition),
- A template is generated for eyes' iris region for biometric identification.
- This template is studied in regard with the other templates stored in a database for comparison until either a matching model is found or no match is detected.
- If a match is recognized, the client is declared identified and acknowledged
- If no match is recognized, the client remains unidentified and anonymous.

CHAPTER-2

OBJECTIVE

The purpose of this project will be to implement an iris recognition and identification system which can authenticate the claimed performance of the methodology. The development tool used will be MATLAB®, and emphasis will be only on the software for exhibiting recognition, and not hardware for capturing an eye image. MATLAB® provides an excellent RAD(Rapid Application Development) environment, with its image processing toolbox, and high level computing techniques. Two sets of eye images from different databases are considered to confirm the certainity of system programming. The two data base being:

 CASIA: a database of 756 greyscale eye images courtesy of The Chinese Academy of Sciences – Institute of Automation, and

 LEI: a database of 120 digital greyscale images courtesy of the Lion's Eye Institute.

CHAPTER-3

IRIS RECOGNITION METHOD

3.1. The Iris Recognition Process

The IR recognition method is described in 4 steps:

 Image Acquisition Obtaining the eye image
 Segmentation To locate the iris region in image
 Normalization To achieve invariance to iris size, position and different degree of iris dilation for matching different iris patterns at later stage
 Feature Encoding & Matching To extract as many discriminating features as possible from the iris and result in an iris signature, or template, containing these features

3.2. Image Acquisition

The image is acquired from an online database of eye images. Two public databases were chosen to perform tests upon:
 the UBIRIS database and
 the CASIA database

The former was selected for utilizing standard equipment, and the latter was selected to provide for a comparison

3.3. Segmentation Technique

The principal of the segmentation technique is to locate the iris region in the eye image. This involves locating the internal borderline between the pupil, the small aperture, and the iris region and the exterior borderline between the iris and the sclera, the white colored part of the eye. In most models, these boundaries, which might not be perfectly circular, are modeled as two un-concentric circles.

Iris, the pigmented region of the eye, can be separated from the sclera, the white area of the eye, but is lighter than the pupil. Segmentation techniques are based on this assumption simplifying the process to a large extent. This variation in intensity is employed to threshold the iris image using upper and lower intensity limits. This thresholded image can be further studied by a circular edge detector determining the edges of sclera with iris and iris with a pupil. As a result, iris region is segmented from the rest of eye image. Although this approach simplifies the edge detection step, but in the way introduces the problem of finding safe threshold levels.

3.4. Normalization

After the segmentation technique is executed, normalization is performed in all studied iris recognition systems to obtain:

- invariance to iris size,
- position and
- different degrees of pupil dilation

when matching different iris patterns at a later stage.

3.5. Feature Extraction

The encoding, or feature extraction, aims to segregate as many refined features as could be allowed from the iris template and results in an iris signature, or trademark indication, containing these segregated features. The principal aim of matching process between two templates is to enhance the contingency of ann accurate match for

CHAPTER-4

IMPLEMENTAION AND PROCEDURE

The evaluation methods of images were performed and studied.

For thresholding, the image is required to be converted to Grayscale.

The image is then transferred to function called thresholding.

Based colour difference of iris from sclera, iris can be segmented using the method based on thresholding.

The small region of connected pixels are removed which are not necessary for operation.

Some part of pixels might have been removed that has left a hole in the image, is compensated to avoid any holes in the image.

This will return the thresholded image to the main program. For segmentation, connected component is calculated for the image. Providing thresholded image as input and using 8 connectivity.

This will create a structure called cc that will store 4 fields:

 Connectivity: already mentioned it to be 8 for the 2D image as it provides more accurate output.
 Image size: It is also fixed while normalizing and resizing to 512x512
 NumObjects: Number of distinct objects or components found in the image.
 PixelIdxList: It is a 1-by-NumObject cell array where the kth element in the cell array is a vector containing linear indices of the pixel in the kth object.

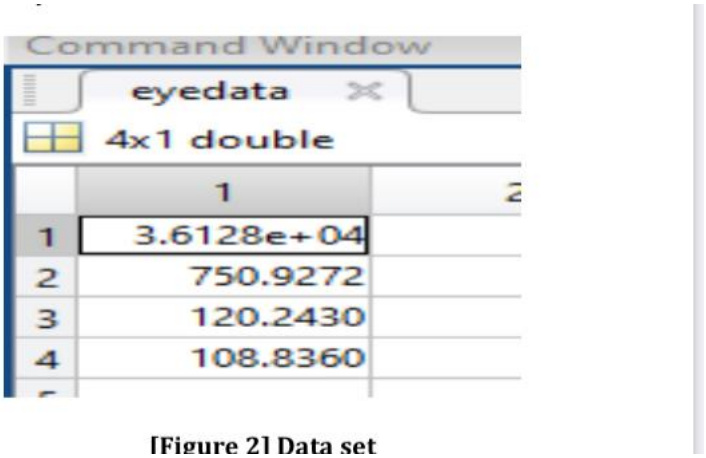
CHATER-5.

EVALUATION

To scrutinize the performance of the iris recognition system, on the whole, tests were performed to locate the best detachment, so that the false match and false acknowledge rate is limited, and to affirm that iris recognition can perform precisely as a biometric for identification of individuals. And additionally affirming that the framework gives precise recognition, the analysis was also supervised. In order to verify the uniqueness of human iris patterns by deducing the number of connected components present in the iris template portrayal.

There are a number of parameters in the iris recognition system, and optimum values for these parameters were required in order to provide the best recognition rate. These parameters include:

1. Connected component: cc The 1-by-NumObect PixelIdxList containing linear indices of the pixel 2. Number of connected components: n 3. Properties of image: k (structure) 4. The mean value of all these data are compiled: eyedata



[Figure 2 [Figure 2] Data set] Data set

All these data provide a means of studying the unique feature of the biometric template, here template being iris.

5.1. Comparison Study:

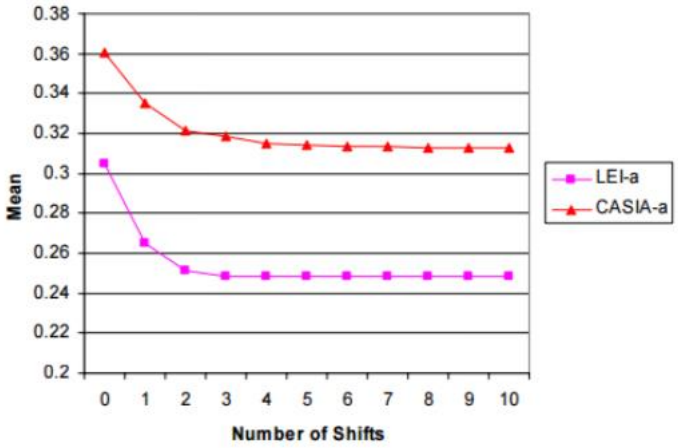
The main aim of an iris recognition system is to have the capacity to accomplish a distinct segregation of intra-class and inter-class Hamming Distance distribution. With clear segregation, a partition Hamming distance value can be picked which enables a choice to be made while contrasting two templates. If the HD between two templates is not as much as the separation point, the templates were created from a similar iris and a match is found. Generally if the HD is more than the separation point the two templates are considered to have been produced from varying sources.

Intra-Class and Inter-Class Hamming Distribution with overlap

For the encoding procedure the yields of each filter ought to be autonomous, so that there are no connections in the encoded layout, or else the filters would be repetitive. For maximum independence, the band-widths of each filter must not cover in the recurrence space, and furthermore the centre frequencies must be spread out.

One element, which will notably influence the identification rate is the radial and angular resolution practiced amid normalization, since this decides the measure of iris pattern information, which goes into encoding the iris layout. The ideal number of template shifts to represent rotational irregularities can be controlled by inspecting the mean and standard deviation of the intra-class distribution. Without template shifting the intra-class Hamming Distance distribution will be all the more arbitrarily distributed, since templates, which are not appropriately aligned, will deliver HD values proportionate to contrasting inter-class templates. As the quantity of shifts increases, the mean of the intra-class distribution will focalize to a constant value, since all rotational irregularities would have been represented for.

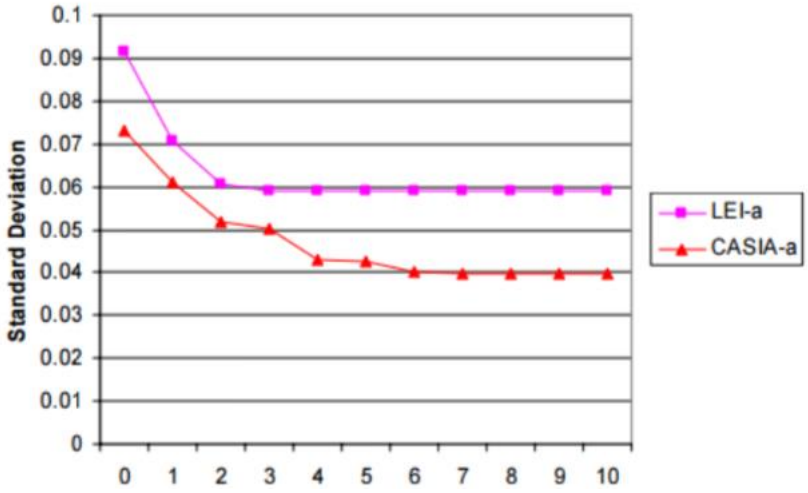
Mean of Intra-class Hamming Distance Distribution vs Number of Shifts



[Figure 3] Mean vs NOS

Mean of the intra-class Hamming distance distribution as a function of the number of shifts performed.

Standard Deviation of Intra-class Hamming Distance Distribution vs Number of shifts



[Figure 4] SD vs NOS

Standard deviation of the intra-class Hamming distance distribution as a function of the number of shifts performed.

5.2. Encountered Issue Depiction

Limitation of imaging the iris is due to the anatomical features of the eye in addition to the noise introduced in the imaging environmental condition. Eyelids together with eyelashes usually congest ad hinder a significant portion of the iris, and this issue must be recognized and tackled in every sturdy iris recognition method. Also, when capturing the picture of eyes under less than perfect conditions, the resolution of the image might be inadequate, and artifacts are unavoidably introduced into the image as noise and blurring due to poor focus.

 Occlusion And Hinderance

The eyelids cover the eye to limit light from going into the eye when required. This is an issue for IR when imaging the eye with visible light, as in the state is while employing standard cameras. The issue can be unraveled by illuminating the eye with light outside the visible range of the spectrum.

Eyelid clogging causes two issues:

 In finding the eye in the image as eyelashes disrupt the circular configuration of the iris region in the image, and
 The eyelid can bring about a substantial portion of the iris pattern to be covered during the template extraction process and hence render it invalid.

 Like the eyelids, eyelashes cause issues in both localization and in the template extraction, although to a lesser degree. Eyelashes are, in contrast with the eyelids, considerably harder to recognize because of their unstructured nature.

 Noise And Disturbances

Iris imaging is a type of assessment, and all the analysis are subjected to faults which can be modeled and handled as disturbances. The noise produced by the imaging sensors and the surrounding electronics is often treated and as white and additive.

 Reflection

The cornea is the outermost transparent portion. This transparent layer protects the eye and admits and helps to focus light waves as they enter the eye. It reflects much of the light is causing a considerable amount of specular reflection. Light sources and surrounding light areas projects on the transparent surface of the eye. These reflections results in in complication in the IR process, clogging the iris pattern and making the location of the eye difficult to estimate as these reflections disfigure the actual structure of the eye.

 Data Loss While Compression

When saving the image data to a file, lossy compression is often used. This introduces information loss and can result in artifacts as visible image blocks and a loss of high-frequency information in the iris pattern.

CHAPTER-6

REVIEW WORKS

Review Works Biometrics studies face, iris, fingerprints, voice, palms, hand geometry, retina, handwriting, gait etc. .
Recognition algorithms requires preprocessing of input image to get better quality of data by tracking various feature points of iris.
Biometric systems captures the feature taking a digital image for iris recognition.
A biometric is characterized by use of a feature that is decidedly unique – so that the chance of any two human having the same features will be minimal .
Person identification based on iris recognition gives one of the most reliable results .
Iris texture features provides a unique high dimensional information that explains why iris recognition based verification has the lowest false acceptance rate among all types of biometric verification systems

CHAPTER-7
RESULT AND DISCUSSION

7.1. Summary Of Result

By using the test results, it can be concluded that an IR system can be constructed using standard equipment, and the performance of such a system would depend on the nature of the images acquired. Regarding the image quality, the light level turned out to be the most important image quality factor followed by focus, reflections, disturbances and level of occlusion and hindrance.

7.2 Future Work

The RAD environment used here can be combined with GUI that is again connected to a database.

The GUI will perform the operation like displaying images and messages accordingly when it is compared to the database. If the image is to identified, the GUI will provide a platform to compare to already stored data in the database. And if it is to be stored, then transfers the data to the database.

The suitable database being MySQL and environment being MATLAB for performing the image processing techniques.

CHAPTER-9

CONCLUSIONS

Iris scanning is a relatively new technology and is incompatible with the very substantial investment that the law enforcement and immigration authorities of some countries have already made into fingerprint recognition. In this paper we highlighted the detection of iris using biotechnology technique.

REFERENCES

1. Casia iris image database. http://www.sinobiometrics.com.

2. C. Barry, N. Ritter. Database of 120 Greyscale Eye Images. Lions Eye Institute, Perth Western Australia.

3. Dal Ho Cho, Kang Ryoung Park, and Dae Woong Rhee. Real-time iris localization for iris recognition in cellular phone. In SNPD, pages 254–259, 2005.

4. LiborMasek. . Recognition of human iris patterns for biometric identification., 2003 http://www.csse.uwa.edu.au/~pk/studentprojects/li bor/.

5. John G. Daugman. How iris recognition works. In Proceed. of 2002 Intern. Confer. on Image Processing, volume 1, 2002.

6. C. Tisse, L. Martin, L. Torres, and M. Robert. Person identification technique using human iris recognition, 2002.

7. J. Daugman. How iris recognition works. Proceedings of 2002 International Conference on Image Processing, Vol. 1, 2002.

8. N. Tun. Recognising Iris Patterns for Person (or Individual) Identification. Honours thesis. The University of Western Australia. 2002.

9. S. Sanderson, J. Erbetta. Authentication for secure environments based on iris scanning technology. IEE Colloquium on Visual Biometrics, 2000

10. R. Wildes. Iris recognition: an emerging biometric technology. Proceedings of the IEEE, Vol. 85, No. 9, 1997

AMC ENGINEERING COLLEGE
DEPARTMENT OF ELECTRICAL AND ELECTRONICS

STATEMENTS
INSTITUTE VISION To be a leader in imparting value based Technical Education and Research for the benefit of society.
INSTITUTE MISSION Provide State Of The Art Infrastructure Facilities. Implement Modern Pedagogical Methods in Delivering the Academic Programs with Experienced and Committed Faculty. Create a Vibrant Ambience that Promotes Learning, Research, Invention and Innovation. Undertake Skill Development Programs for Academic Institutions and Industries. Enhance Institute Industry Interaction Through Collaborative Research and Consultancy. Relentlessly Pursue Professional Excellence with Ethical and Moral Values.
DEPARTMENT VISION
Be a premier department in the field of Electrical and Electronics Engineering for ever changing sustainable needs of the Society.
DEPARTMENT MISSION provide the State-of-the-Art Infrastructure facilities. train in modern tools and techniques in Emerging Technologies. acquire Leadership skills, Professional skills and Ethical values. establish Industry Institute Interaction and make students ready for the Industrial environment. promote co-curricular and extracurricular activities to enhance the overall growth.
PROGRAM EDUCATIONAL OBJECTIVES (PEO)

PEO1	Apply their knowledge and skills of Electrical and Electronics Engineering to solve complex problems in Industry / Government Organization and to pursue higher education & research

PEO2	Work as effective individual or in team with professionalism, ethical values and social responsibilities to manage multi-disciplinary projects.
PEO3	Update their knowledge continuously through lifelong learning to excel in their career
PROGRAM OUTCOMES (PO)	
PO1	**Engineering knowledge**: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
PO2	**Problem analysis**: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.
PO3	**Design/development of solutions**: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
PO4	**Conduct investigations of complex problems**: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
PO5	**Modern tool usage**: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.
PO6	**The engineer and society**: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.
PO7	**Environment and sustainability**: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.
PO8	**Ethics**: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
PO9	**Individual and team work**: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
PO10	**Communication**: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
PO11	**Project management and finance**: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.
PO12	**Life-long learning**: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.
PROGRAM SPECIFIC OUTCOMES (PSO)	
PSO1	The graduate will be able to apply the knowledge acquired from strong fundamentals of Mathematics, Science and Engineering Subjects to identify, formulate, design and investigate complex Engineering Problems of Electrical and Electronics to pursue a successful career/higher studies.
PSO2	Be a professional to apply appropriate techniques and Modern Engineering Software tools to design and develop Electrical Systems, also engage in lifelong learning and successfully acquire leadership qualities, communication skills, ethical attitude, achieve competence to excel individually, work efficiently in a team and become an entrepreneur.

COURSE INFORMATION SHEET	
PROGRAMME DETAILS	
Name of the Programme	B.E
Branch	EEE
COURSE DETAILS	
Course Title	SEMINAR
Course Code	15EEP78&15EEP85
Year	4
Sem/Sec	7&8
Academic year	2019-2020
Batch	2014,2015, 2016,2017
Number of Students	04
GUIDE DETAILS	
Name & Designation of the faculty	SHIVALINGASWAMY G D
Department	EEE

COURSE OUTCOMES (CO)- PEO/PO/PSO Mapping

CO.No.	OUTCOMES	**Bloom's Cognitive Level**	PO	**PSO**
CO1	Able to generate ,develop idea and information to carry out project work	Analyze	1,2,3,4,5,6,7,12	1,2
CO2	Able to Identify a real-life problems	Analyze	1,2,3,4,5,6,7,12	1,2
CO3	Able to adapt skills to communicate effectively	Apply	8,9,10	
CO4	Able to adapt collaborative skills to work in team	Apply	8,9	1,2
CO5	Able to Analyze and Implement a tangible solution using available resources	Evaluate/Create	1,2,3,4	1,2

Strength of CO Mapping to PEO/PO/PSOs with Justification:

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
CO1	3	3	3	1	2	1	1					1	3	3
CO2	3	3	3	2	2	1	1					1	3	3
CO3								1	3	1				
CO4								1	3				1	1
CO5	3	3	3	1									3	3

3 – strong; 2 – medium; 1 – weak

CO-PEO/PO/PSO	Justification
CO1->PO1(3)	pply the knowledge engineering fundamentals, to generate ,develop idea and information to carry out project work . Hence a strong mapping
CO1->PO2(3)	Analyze the complex problem in developing idea and information to carry out project work. Hence a strong mapping.
CO1->PO3(3)	le to, develop idea and information to carry out project work . Hence a strong mapping. Hence a strong mapping.
CO1->PO4(1)	Able to use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions. Hence weak mapping.
CO1-> PO5(2)	Able to apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities, to generate ,develop idea and information to carry out project work Hence a moderate mapping.
CO1-> PO6(1)	Able to apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice. Hence a weak mapping.
CO1-> PO7(1)	Able to Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.. Hence a weak mapping
CO1-> PO12(1)	Able to engage in independent and life-long learning in the broadest context of technological change.. Hence aweak mapping.
CO2-> PO1(3)	Able to Apply the knowledge engineering fundamentals to Identify a real-life problems , Hence a strong mapping
CO2-> PO2(3)	Able to Identify a real-life problems, Hence a strong mapping .
CO2-> PO3(3)	Able to design solution for real life technical problems.. Hence a strong mapping.
CO2-> PO4(2)	Able to use research based Knowledge to Identify a real-life problems. Hence moderate mapping.
CO2-> PO5(2)	Able to Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools. to Identify a real-life problems. Hence moderate mapping.
CO2-> PO6(1)	Able to Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues to Identify a real-life problems Hence a weak mapping
CO2-> PO7(1)	Able to understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development. Hence weak mapping
CO2-> PO12(1)	Able to engage in independent and life-long learning in the broadest context of technological change. Hence a weak mapping
CO3-> PO8(1)	Able to apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice to adapt skills to communicate effectively . Hence a weak mapping.
CO3-> PO9(3)	Able to Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary. Hence a strong mapping
CO3-> PO10(1)	Able to Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions. Hence a weak mapping.
CO4-> PO8(1)	Able to adapt collaborative skills to work in team. Hence a weak mapping.
	Apply to Function effectively as an individual, and as a member or leader in diverse

CO4-> PO9(3)	teams, and in multidisciplinary settings . Hence a strong mapping
CO5-> PO1(3)	Able to implement a tangible solution using the knowledge of mathematics, science, engineering fundamentals. Hence a strong mapping.
CO5-> PO2(3)	Able to Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences. Hence a strong mapping.
CO5-> PO3(3)	Able to Design solutions for complex engineering problems and design system components or processes that meet the specified needs. Hence a strong mapping.
CO5-> PO4(1)	Able to use research-based knowledge and research methods to implement a tangible solution using available resources. Hence a weak mapping.

1.

VISVESVARAYA TECHNOLOGICAL UNIVERSITY

Jnana Sangama, Belgavi - 590018

2019-20

A Technical Seminar Report on

SEISMIC RETROFITTING OF BUILDING

Submitted in partial fulfillment of the requirement for the award of degree of

BACHELOR OF ENGINEERING IN CIVIL ENGINEERING

Submitted by

AKASH K ANIL
[USN: 1AM16CV003]

Under the Guidance of

Mr. GANESH S. S. Y,
Assistant Professor

AMC ENGINEERING COLLEGE

DEPARTMENT OF CIVIL ENGINEERING

[NBA & NAAC Accredited, affiliated to VTU, Belagavi, Approved by AICTE, New Delhi]

18th K.M. Bannerghatta Main Road Bangalore, 560083

AMC ENGINEERING COLLEGE

[NBA & NAAC Accredited, affiliated to VTU, Belagavi, Approved by AICTE, New Delhi]

18th K.M. Bannerghatta Main Road Bangalore,560083

DEPARTMENT OF CIVIL ENGINEERING

CERTIFICATE

This is to certify that the evaluation of technical seminar report titled "**SEISMIC RETROFITTING OF BUILDING**" is a bonafide work carried out by **AKASH K ANIL (1AM16CV003),** in partial fulfillment for the award of "Bachelor of Engineering" in Civil Engineering of the Visvesvaraya Technological University, Belagavi during the year 2019-20. It is certified that all corrections/suggestions indicated for the internal assessment have been incorporated in the report deposited in the department library. The technical seminar report has been approved as it satisfies the academic requirements in respect of the technical seminar work prescribed for the said degree.

Signature of Guide	**Signature of HOD**	**Signature of Principal**
Mr. Ganesh S. S. Y	Dr. Shashishankar. A	Dr. A. G. Nataraj

DECLARATION

I, Akash K Anil, student of eighth semester B. E in the Department of Civil Engineering, AMC Engineering College Bengaluru, declare that the technical seminar entitled **"SEISMIC RETROFITTING OF BUILDING**" has been carried out by me and submitted in partial fulfillment of the course requirements for the award of degree in Bachelor of Engineering in Civil Engineering of Visvesvaraya Technological University, Belagavi during the academic year 2019-20.

Date: **AKASH K ANIL**

Place: Bengaluru **(USN:1AM16CV003)**

ACKNOWLEDGEMENT

Any achievement, be it scholastic or otherwise does not depend solely on the individual efforts but on the guidance, encouragement and cooperation of intellectuals, elders and friends. Several personalities, in their own capacities have helped me in carrying out this technical seminar. I would like to take this opportunity to thank them all.

I wish to express my profound gratitude to **Dr. K. R. Paramahamsa**, AMCEC, Bengaluru for his moral support towards completion of my technical seminar.

I wish to express my profound gratitude to **Dr. A.G Nataraj**, Principal, AMCEC, Bengaluru for his moral support towards completion of my technical seminar.

I would like to express my deepest gratitude to **Dr. Shashishankar A**, Head of Department, Civil Engineering, AMCEC, Bengaluru, for his continuous support and encouragement.

I would like to acknowledge the unbridled enthusiasm of my technical seminar guide

Mr. Ganesh S. S. Y, Assistant Professor, Department of Civil Engineering, AMCEC, Bengaluru, for his encouragement and valuable guidance throughout my technical seminar.

I would like to acknowledge the unbridled enthusiasm of my technical seminar coordinator, **Mr. Ravithej,** Assistant professor, Department of Civil Engineering, AMCEC, Bengaluru, for her encouragement and valuable guidance throughout my technical seminar.

I thank my Parents, and all the teaching and non-teaching faculty members of Department of Civil Engineering for their constant support and encouragement. Last, but not the least, I would like to thank my friends who provided me with valuable suggestions to my work during technical seminar.

INTRODUCTION

Earthquakes around the world are single-handedly responsible for the destruction to life and property in large numbers.

An earthquake is the result of a sudden release of energy in the earth's crust that creates seismic waves. The seismic activity of an area refers to the frequency, type and size of earthquakes experienced over a period.

In order to mitigate such hazards, it is important to incorporate norms that will enhance the seismic performance of structures.

Seismic Retrofitting is a collection of mitigation technique for Earthquake engineering. Seismic Retrofitting Techniques are required for concrete constructions which are vulnerable to damage and failures by seismic forces.

SEISMIC RETROFITTING

- Seismic, by the word itself means something related to earthquake or vibrations of the earth and its crust. It happens due to some disturbances below the earth's crust which cannot be seen.
- Seismic retrofitting is the modification of existing structures to make them more resistant to seismic activity, ground motion, or soil failure due to earthquakes.
- Seismic retrofitting of vulnerable structures is critical to reducing risk. It is important for protecting the lives and assets of building occupants and the continuity of their work.
- If people live or work in retrofitted structures, they are less likely to be injured during an earthquake.

Figure: SEISMIC RETROFITTING

AIMS and OBJECTIVES

- **PUBLIC SAFETY**: The goal is to protect human life, ensuring that the structure will not collapse upon its occupants or passer-by.
- **STRUCTURE SURVIVABILITY**: The goal is that the structure, while remaining safe may require extensive repair but not replacement.
- **STRUCTURE FUNCTIONALITY**: Primary structure undamaged and the structure is undiminished in utility for its primary application.
- **STRUCTURE UNAFFECTED**: A high level of retrofit is preferred for historic structures of cultural significance

NECESSITY FOR SEISMIC RETROFITTING IN EXISTING BUILDINGS

The need for seismic retrofitting in existing buildings can arise due to the following reasons:

- Buildings not designed according to the codes of practice.
- Deterioration of strength of the buildings.
- Not considering the safety of buildings while construction.

RAPID VISUAL SCREENING (RVS)

RAPID VISUAL SCREENING (RVS) is an important step to be taken prior to retrofitting. RVS is a method to estimate the seismic vulnerability. It is designed to be implemented without performing any structural calculations.

The procedure utilises a damageability grading system that requires the evaluator to perform the following:

- Identify the primary structural lateral load-resisting system,
- Identify building attributes that modify the seismic performance expected for this lateral load-resisting system along with non-structural components.

The inspection, data collection and decision-making process typically occurs at the building site and is expected to take couple of hours for a building, depending on its size.

SEISMIC RETROFITTING TECHNIQUES

Common seismic retrofitting techniques are as follows:

- Base Isolators
- Supplementary dampers
- Tuned mass dampers
- Slosh tank
- Active control system

BASE ISOLATORS

Base Isolation is a technique developed to prevent or minimize damage to buildings during an earthquake. Placing flexible isolation systems between the foundation and the superstructure.

- It has been used in New Zealand, India, Japan, Italy and the USA.
- When a building is built-away (isolated) from the ground, resting on flexible bearings or pads known as base isolators, it will move little or not at all during an earthquake.

Base Isolators are constructed using the following basic components which are generally placed in layers:

- Rubber: It provides flexibility at the end of an earthquake; the rubber bearings will slowly bring the building back to its original position which takes months to happen.
- Lead plug: it has the plastic property during an earthquake, the kinetic energy of the earthquake is absorbed into heat energy as the lead is deformed.

- Steel: If layers of steel are used with rubber, the bearings can move in the horizontal direction but is stiff in the vertical direction.

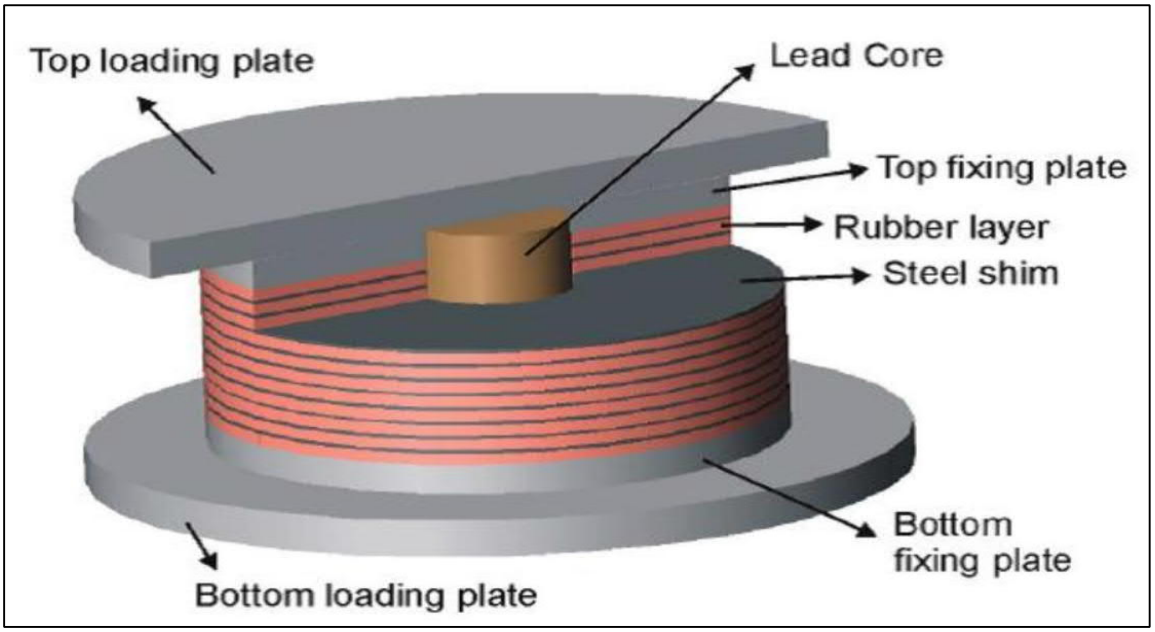


Figure: BASE ISOLATORS

SUPPLEMENTARY DAMPERS

- A Supplementary Damping System (SDS) is the most efficient and cost-effective way to achieve energy dissipation system that is incorporated into the design of a structure to absorb vibration energy, thereby reducing motion.
- This would inadvertently mean decreasing the energy dissipation demand on the structural components i.e. beams/columns/slabs thereby increasing the survivability of the building structure.
- Dampers are mechanical devices that look somewhat like huge shock absorbers and their function is to absorb and dissipate the energy supplied by the ground movement during earthquake.
- The energy absorbed by the dampers gets converted into heat which is then dissipated harmlessly into atmosphere.
- Dampers thus work to absorb earthquake shocks ensuring that the structural members (beams and columns) remain unharmed.

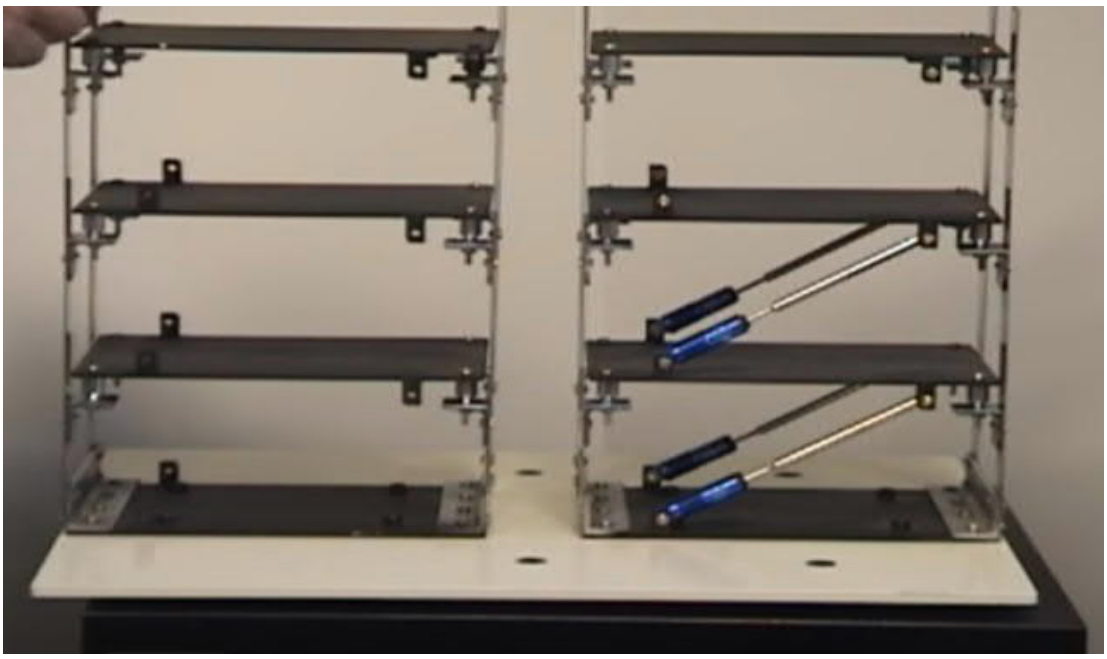


Figure: Supplementary Dampers

TUNED MASS DAMPERS

- A tuned mass damper, also known as a harmonic absorber, is a device mounted in structures to reduce the amplitude of mechanical vibrations.
- Tuned mass dampers stabilize against violent motion caused by harmonic vibration.
- Their application can prevent discomfort, damage or outright structural failure.
- They are frequently used in power transmission, automobiles and buildings.

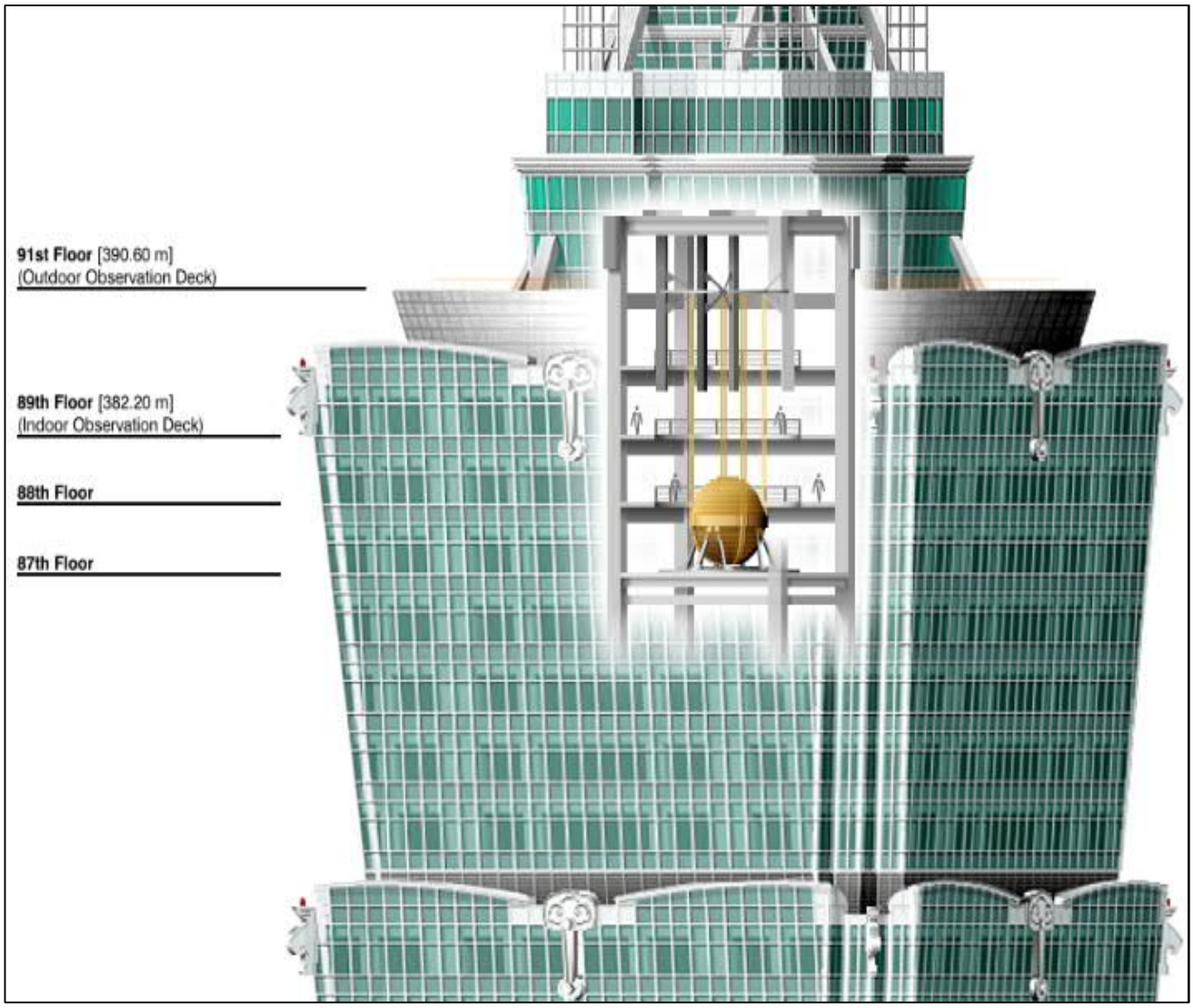
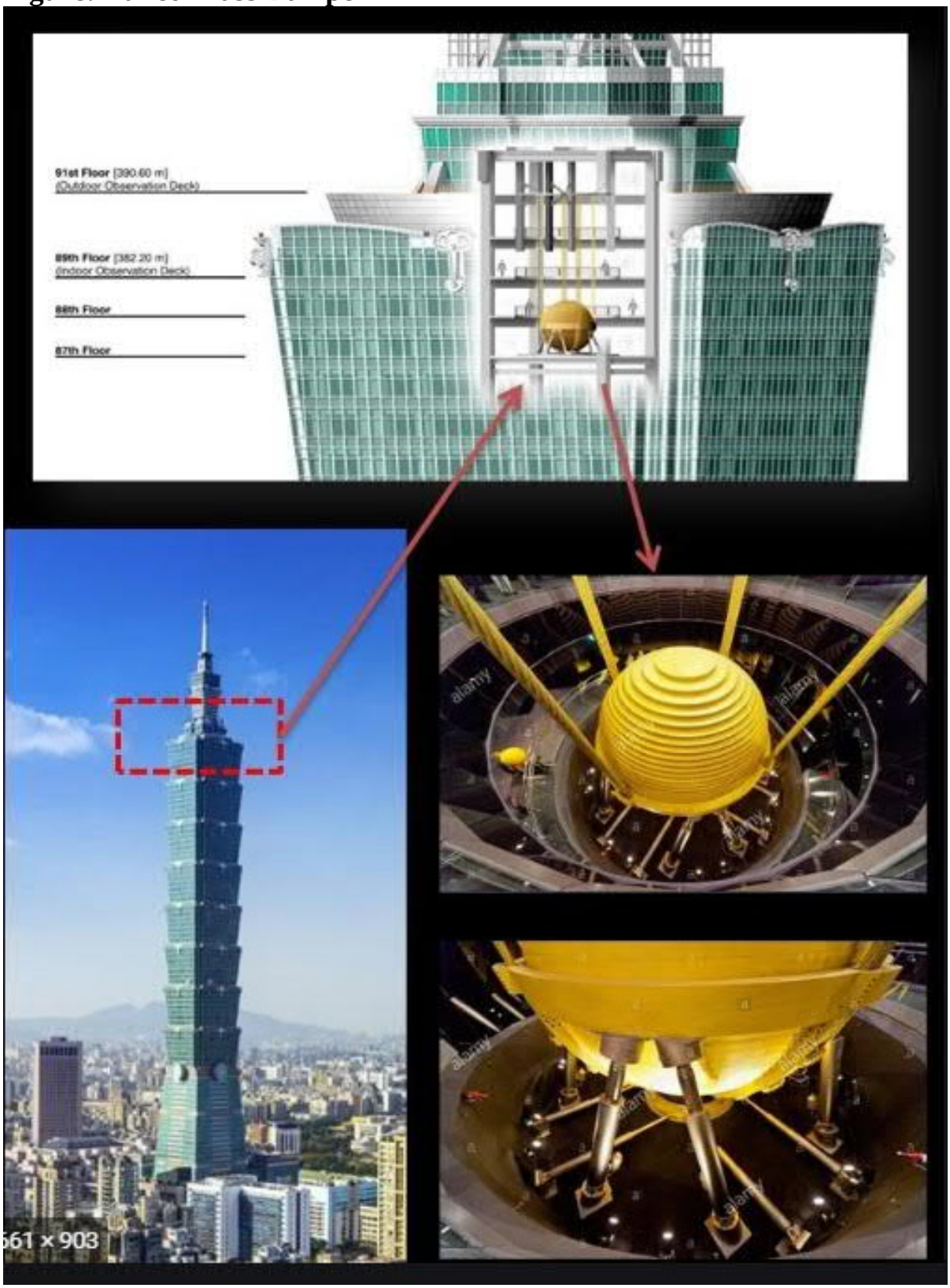


Figure: Tuned Mass Damper



SLOSH TANK

- In fluid dynamics, slosh refers to the movement of liquid inside another object undergoing motion. Nowadays, one of the biggest challenges that engineering faces is to reduce structure motion due to external loadings especially in high rise buildings.

- Slosh tank is one of the inventions that can be installed in different locations and levels into a structure, in order to increase dampening (energy absorbing mechanism) and decrease vibrations.

- It can either be installed on the top floor of a structure or in some certain floors or even at each floor of a building.

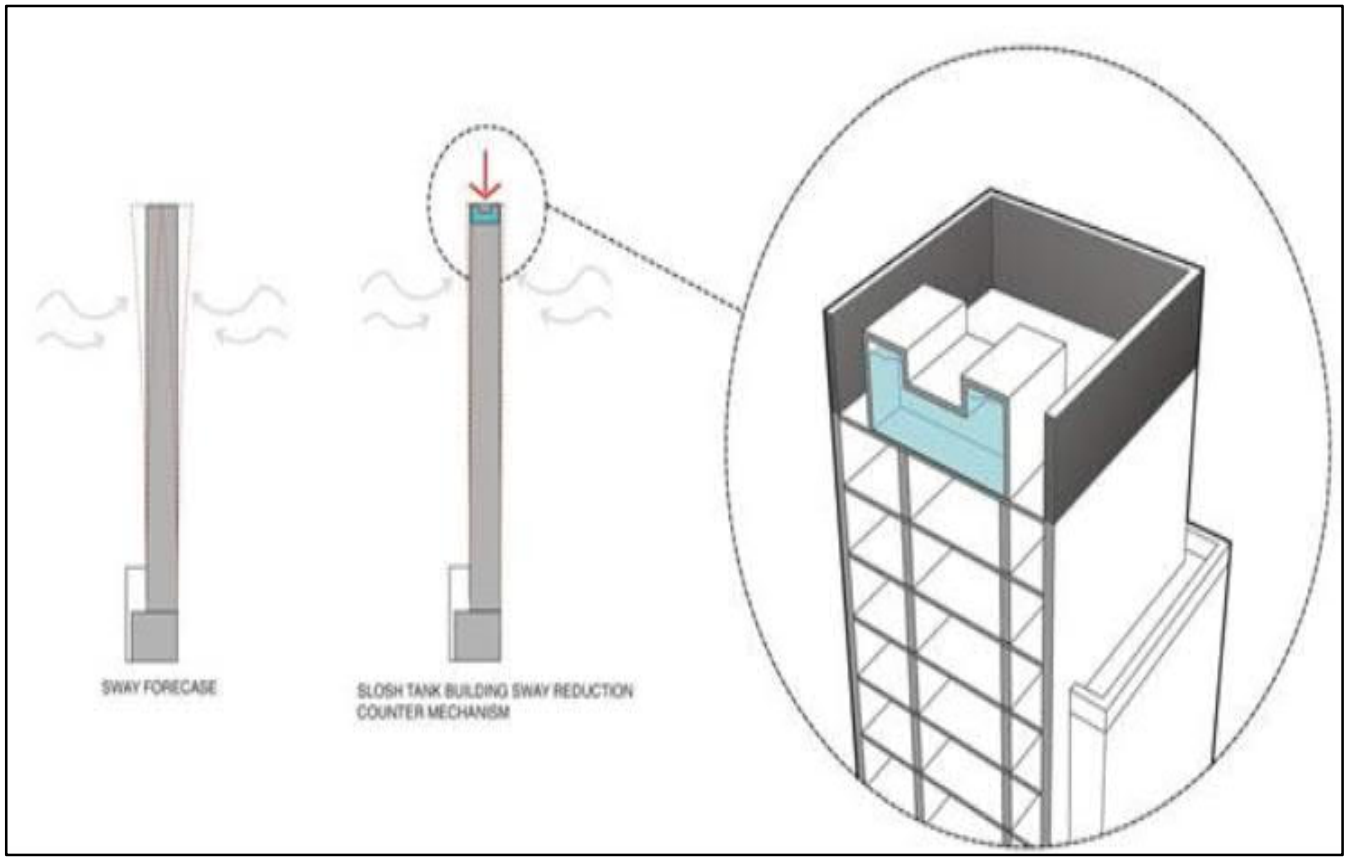


Figure: Slosh Tank

ACTIVE CONTROL SYSTEM

- The use of active control systems and some combinations of active and passive systems, so called hybrid systems, as a means of structural protection against seismic loads has received considerable attention in recent years.

- Active/hybrid control systems are force delivery devices integrated with real-time processing evaluators/controllers and sensors within the structure.

- An active structural control system consists of the following :
  - Sensor located about the structure to measure either external excitations, or structural response variables, or both.
  - Devices to process the measured information and to compute necessary control force needed based on a given control algorithm.

Actuators, usually powered by external sources, to produce the required forces

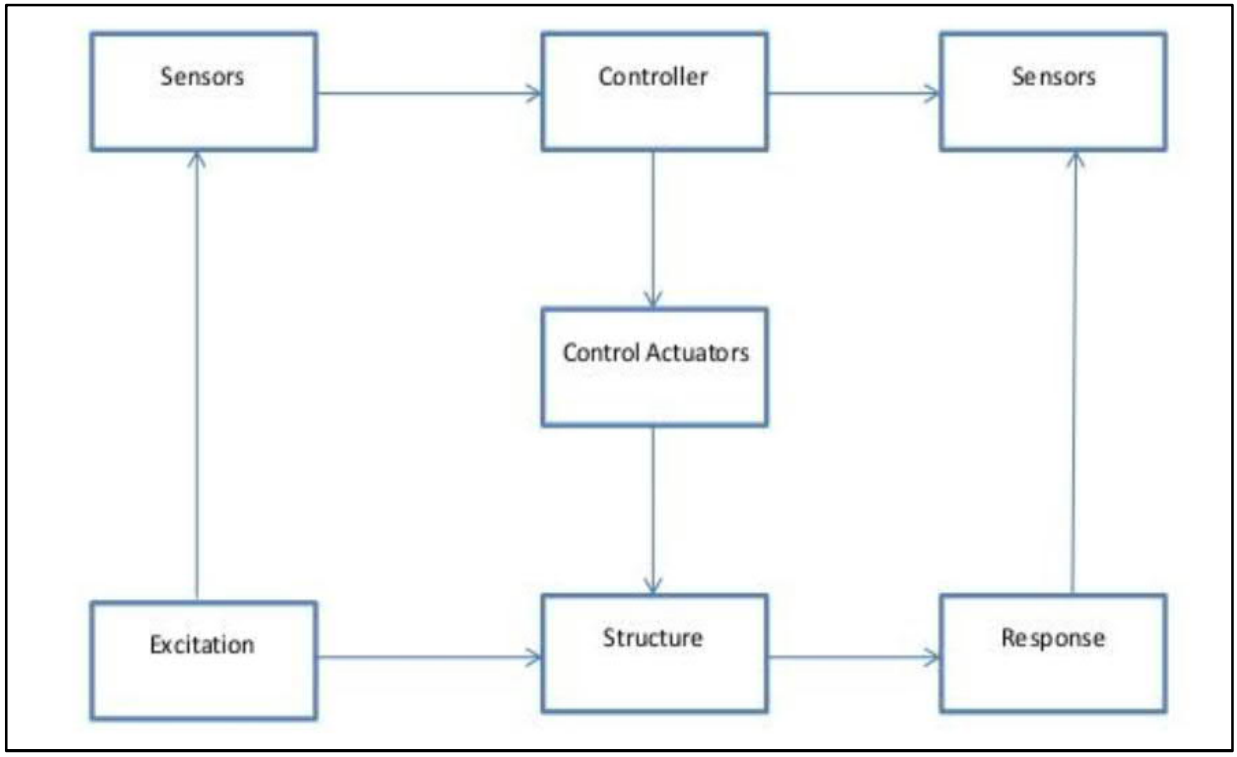


Figure: Schematic representation of Active Control Systems

ADVANTAGES AND DISADVANTAGES

ADVANTAGES

- Building can remain serviceable throughout construction
- They do not depend on an external power source for their operation
- They can also be introduced in upgrading structures
- They require low maintenance
- Their properties can be adjusted in the field

DISADVANTAGES

- Challenging to implement in an efficient manner
- Not suitable for buildings rested on soft soil
- It will be expensive
- A large mass or a large space is needed for their installation
- Inefficient for high rise buildings

IS CODES FOR SEISMIC DESIGN (REFERENCE)

LIST OF IS (INDIAN STANDARDS) CODES REQUIRED FOR SEISMIC DESIGN ARE AS FOLLOWS:

- IS 4326:1993 - Earthquake resistant design
- IS 13827:1993 - Earthquake resistance of earthen buildings
- IS 13828:1993 - Earthquake resistance of low strength masonry buildings
- IS 13920:1993 - Ductile detailing of reinforced concrete structures
- IS 13935 - Seismic strengthening of buildings
- IS 1893 - Earthquake resistant design of structures

VISVESVARAYA TECHNOLOGICAL UNIVERSITY

"Jnana Sangama", Belagavi 590 018

Technical Seminar report on

"DIGITAL TWIN SPARK IGNITION SYSTEM"

Submitted in partial fulfillment for the award of degree of

BACHELOR OF ENGINEERING

in

MECHANICAL ENGINEERING

by

JNANESHWAR M POOJERI

(1AM16ME059)

Under the Guidance of

DR. SHANTHALA K.

Associate Professor

Department of Mechanical Engineering

AMC Engineering College, Bengaluru - 560 083.

Department of Mechanical Engineering

AMC ENGINEERING COLLEGE,

18th K.M. Bannerghatta main road, Bengaluru – 560 083

2019 – 2020

AMC ENGINEERING COLLEGE, BENGALURU-560 083

DEPARTMENT OF MECHANICAL ENGINEERING

CERTIFICATE

Certified that the Technical Seminar Report entitled **"Digital Twin Spark Ignition System"** presented by **Mr. JNANESHWAR M POOJERI** bearing USN **1AM16ME059,** bonafide Student of **Mechanical Engineering** in **AMC Engineering College** of the **Visveswaraya Technological University, Belagavi** during the year **2019-2020.** It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated. The report has been approved as it satisfies the partial fulfillment of the course requirements for the award of degree of Bachelor of Engineering in Mechanical Engineering of Visveswaraya Technological University, Belagavi.

Signature Guide	Signature HOD	Signature Principal

AMC ENGINEERING COLLEGE, BENGALURU-560 083

DEPARTMENT OF MECHANICAL ENGINEERING

DECLARATION

I, **JNANESHWAR M POOJERI ,** student of 8th semester B.E.,Mechanical Engineering in AMC Engineering College, hereby declare that the Technical Seminar Report entitled **"Digital Twin Spark Ignition System"** has been presented by me at AMC Engineering College ,Bengaluru and submitted in partial fulfillment of the course requirements for the award of degree of **Bachelor of Engineering in Mechanical Engineering of Visveswaraya Technological University, Belagavi,** during the academic year **2019-2020 .** I also declare that, to the best of my knowledge and belief, the work reported here does not from part of any other dissertation on the basis of which a degree or award was confirmed on an earlier occasion on this by any other student.

JNANESHWAR M POOJERI
1AM16ME059

ACKNOWLEDGEMENT

I thank the Almighty for his presence and guidance throughout every venture of my life and for showering me with His blessings.

Words are short for expressing my deepest sense of gratitude and sincere thanks to my guide **Dr. Shanthala K., AMC Engineering College, Bengaluru**. She has been my greatest source of inspiration right from instilling the very idea of this research work into my mind. From the beginning till now he has provided me valuable help, guidance, constant encouragement and support in all the stages of the research.

I take this opportunity to thank **Dr. A G Nataraj, Principal and Head of the Research Centre, AMC Engineering College, Bengaluru** for providing valuable suggestions and opportunity to undergo this research.

I am grateful to **Dr. Anantha Keshava Murthy, Professor and Dean of Academics, AMC Engineering College, Bengaluru,** for his support and for providing necessary facilities to carry out the research.

I am grateful to **Dr. Girisha. C, Professor and Head, Department of Mechanical Engineering** for his support and for providing necessary facilities to carry out the research.

I am indebted to my friends, all the faculty and staff members of the department for providing me with the relevant information and helped me in different capacities in carrying out this Report.

JNANESHWAR M POOJERI 1AM16ME059

ABSTRACT

The latest trend of new generation bikes and cars is to use new technology and high speed. It has become a fashion for the people especially living in urban areas to ride such vehicles. Now the companies even want to launch such vehicles that attract the younger generation. This can be achieved by technology known as DTS-i (digital twin spark ignition) system which combines strong performance and fuel efficiency. The improved engine efficiency modes have also resulted in lower fuel consumption. DTS-i systems meet the Government of India's emission norms for 2005 right. Spark ignition is one of the most vital systems of a petrol engine. Any variation in the spark timing and number of sparks per minute affects the engine performance severely. Thus, a good design and control of the system parameters becomes most essential for optimum performance of an engine. Due to Digital Twin Spark Ignition System it is possible to combine strong performance and higher fuel efficiency. DTS-i offers many advantages over conventional mechanical spark ignition system.

CONTENTS

1. INTRODUCTION

Rapid combustion is the basic requirement for knock free operation of an S.I engine. The important attributes of rapid combustion are improved tradeoff between efficiency and NOX emissions, greater tolerance towards EGR (exhaust gas recirculation) or excess air, which can improve vehicle drivability and greater knock resistance, thereby allowing fuel economy with higher compression ratios.

Multiple ignition system is one of the techniques to achieve rapid combustion. Multiple spark plug engines often use the initiation of flame propagation at two or more number of points in the combustion chamber depending on the number of spark plugs employed. If two plugs are employed the flame front travels from two points in the cylinder and the effective distance to be travelled by each flame is reduced. The concept of dual plug spark ignition is under consideration for more than three decades. Several experimental studies were made in the area of dual ignition engines regarding optimization of spark plug location and to prove their efficient operation at part loads, extended EGR tolerance, extended lean misfire limit and relatively clean burning compared with single spark ignition systems.

2. DIGITAL TWIN SPARK IGNITION SYSTEM

DTS-i stands for Digital Twin Spark Ignition. Bajaj developed a few years ago and has incorporated in many of its current engines, takes care of the slow rate of combustion in a simple but novel way. The cylinder head is equipped with two spark plugs, instead of the usual one. By generating two sparks at either ends of the combustion chamber, (approximately 90° to the valve axis) the air-fuel mixture gets ignited in a way that creates two flame fronts and, therefore, a reduction in flame travel of the order of 40 per cent is achieved. A fast rate of combustion is achieved leading to faster rise in pressure. The obvious outcome of this is more torque, better fuel efficiency and lower emissions.

What does the Digital, Twin, Spark Ignition means?

- ✓ Digital - Since the spark generation will be initiated by a microchip.
- ✓ Twin - Since two spark plugs will be used.
- ✓ Spark-ignition - Since the ignition will be done via a spark.

3. MAIN CHARACTERISTICS

- Digital electronic ignition with two plugs per cylinder and two ignition distributors.
- Ignition with a Digital C.D.I.
- Injection fuel feed with integrated electronic twin spark ignition.
- A high specific power.
- Compact design and Superior balance.

4.HISTORY

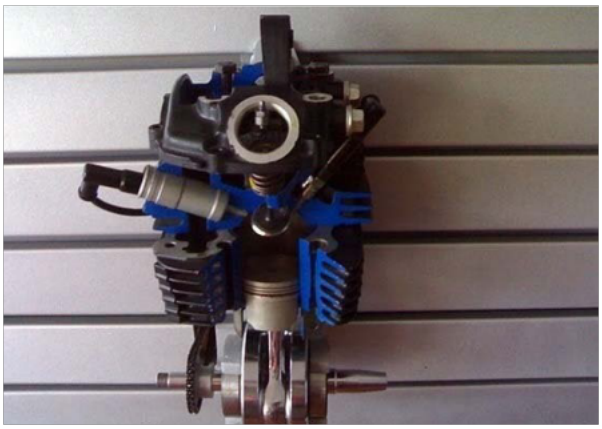
DTS-i stands for Digital Twin Spark Ignition, a Bajaj Auto trademark. Bajaj Auto holds an Indian patent for the DTS-i technology. The Alfa Romeo Twin-Spark engines, the BMW F650 Funduro which was sold in India from 1995 to 1997 also had a twin spark plug technology, and the Rotax motorcycle engines, more recently Honda's iDSI Vehicle engines use a similar arrangement of two spark-plugs. However very few small capacity engines did eventually implement such a scheme in their production prototypes.

DIFFERENCE BETWEEN SINGLE SPARK PLUG AND TWIN SPARK PLUG

CHARACTERISTICS	SINGLE SPARK PLUG (Hero Honda CBZ)	TWIN SPARK PLUG (Pulsar 180 cc)
1. **Power**	12.08PS at 8000rpm	16PS at 8000rpm
2. **Torque**	12.03Nm at 6500rpm	14.72Nm at6500rpm
3. **Speed**	100 kmph	118 kmph
4. **Mileage**	50-55 kmpl	50-55 kmpl
5. **Fuel Control System**	Transient power fuel control	Electronic Control

5. CONSTRUCTION OF DTS-I ENGINE

Digital spark technology is currently used in Bajaj motor cycles in India, because they have the patent right. Digital twin spark ignition technology powered engine has two spark plugs. It is located at opposite sides of combustion chamber. This DTS-I technology will have greater combustion rate because of twin spark plug located around it. The engine combust fuel at double rate than normal. This enhances both engine life and fuel efficiency. It is mapped by the digital electronic control box which also handles fuel ignition and valve timing.



Microprocessors continuously senses speed and load of the engine and respond by altering the ignition timing there by optimizing power and fuel economy.

The main components of DTS-i engine

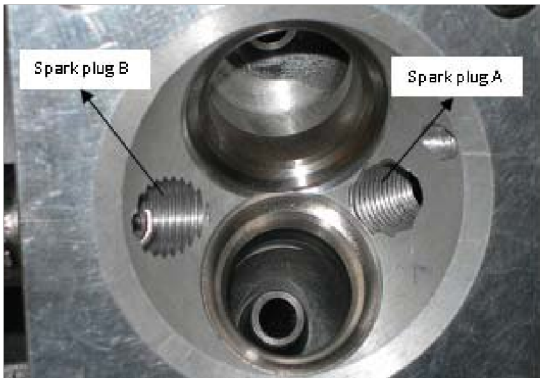
5.1: - Cylinder

5.2: - Crank Shaft
5.3: - Connecting rod
5.4: - Fly wheel
5.5: - Carburetor
5.6: - 2-sparkplug
5.7: - 2-ports & 2- Valves

5.1 Cylinder: -

A cylinder is the central working part of a reciprocating engine, the space in which a piston travel. Which is typically cast from aluminum or cast-iron before receiving precision machine work. Cylinders may be sleeved (lined with a harder metal) or sleeveless (with a wear-resistant coating such as Nikasil).

A piston is seated inside each cylinder by several metal piston rings fitted around its outside surface in machined grooves; typically, two for compressional sealing and one to seal the oil. The rings make near contact with the cylinder walls (sleeved or sleeveless), riding on a thin layer of lubricating oil; essential to keep the engine from seizing and necessitating a cylinder wall's durable surface.



Cylinder block this is a cast structure with cylindrical holes bored to guide and support the pistons and to harness the working gases. Cylinder head is casting encloses the combustion end of the cylinder block and houses both the inlet and exhaust poppet valves and their ports admit air – fuel mixture and to exhaust the combustion products.

to

5.2 Crank shaft: -

The crankshaft, sometimes abbreviated to crank, is responsible for conversion between reciprocating motion and rotational motion. In a reciprocating engine, it translates reciprocating linear piston motion into rotational motion. In order to do the conversion between two motions, the crankshaft has "crank throws" or "crankpins", additional bearing surfaces whose axis is offset from that of the crank, to which the "big ends" of the connecting rods from cylinder attach.

It is typically connected to a flywheel to reduce the pulsation characteristics of the four – stroke cycle, and sometimes the torsional or vibrational damper at the opposite end, to reduce the torsional vibrations often Caused along the length of the crankshaft by the cylinder farthest from the output end acting on the torsional elasticity of the metal.

5.3 Connecting Rod: -

In a reciprocating piston engine, the connecting rod connects the piston to the crank or crankshaft. Together with the crank, they form a simple mechanism that converts reciprocating motion into rotating motion.

As a connecting rod is rigid, it may transmit either a push or a pull and so the rod may rotate the crank through both halves of a revolution, i.e. piston pushing and piston pulling. Earlier mechanisms, such as chains, could only pull. In a few two-stroke engines, the connecting rod is only required to push.

In modern automotive internal combustion engines, the connecting rods are most usually made of steel for production engines, but can be made of T6-2024 and T651-7075 aluminum alloys (for lightness and the ability to absorb high impact at the expense of durability) or titanium (for a combination of lightness with strength, at higher cost) for high performance engines, or of cast iron for applications such as motor scooters. They are not rigidly fixed at either end, so that the angle between the connecting rod and the piston can change as the rod moves up and down and rotates around the crankshaft. Connecting rods, especially in racing engines, may be called "billet" rods, if they are machined out of a solid billet of metal, rather than being cast or forged.

5.4 Flywheel: -

A flywheel is a rotating mechanical device that is used to store rotational energy. Flywheels are often used to provide continuous energy in systems where the energy source is not continuous. In such cases, the flywheel stores energy when torque is applied by the energy source, and it releases stored energy when the energy source is not applying torque to it. A flywheel is used to maintain constant angular velocity of the crankshaft in a reciprocating engine.

The amount of energy stored in a flywheel is proportional to the square of its

rotational speed. Energy is transferred to a flywheel by applying torque to it, thereby increasing its rotational speed, and hence its stored energy. Conversely, a flywheel releases stored energy by applying torque to a mechanical load, thereby decreasing its rotational speed.

Flywheels are typically made of steel and rotate on conventional bearings; these are generally limited to a revolution rate of a few thousand rpm. Some modern flywheels are made of carbon fiber materials and employ magnetic bearings, enabling them to revolve at speeds up to 60,000 rpm.

5.5 Carburetor: -

Spark ignition engines normally use volatile liquid fuels. Preparation of fuel-air mixture is done outside the engine cylinder and formation of a homogeneous mixture is normally not completed in the inlet manifold. Fuel droplets which remain in suspension continue to evaporate and mix with air even during suction and compression processes. The process of mixture preparation is extremely important for spark ignition engines. The purpose of

carburetion is to provide a combustible mixture of fuel and air in the required quantity and quality for efficient operation of the engines under all conditions.

The process of information of a combustible fuel-air mixture by mixing the proper amount of fuel with air before admission to engine cylinder is called carburetion and the device which does this job is called a carburetor.

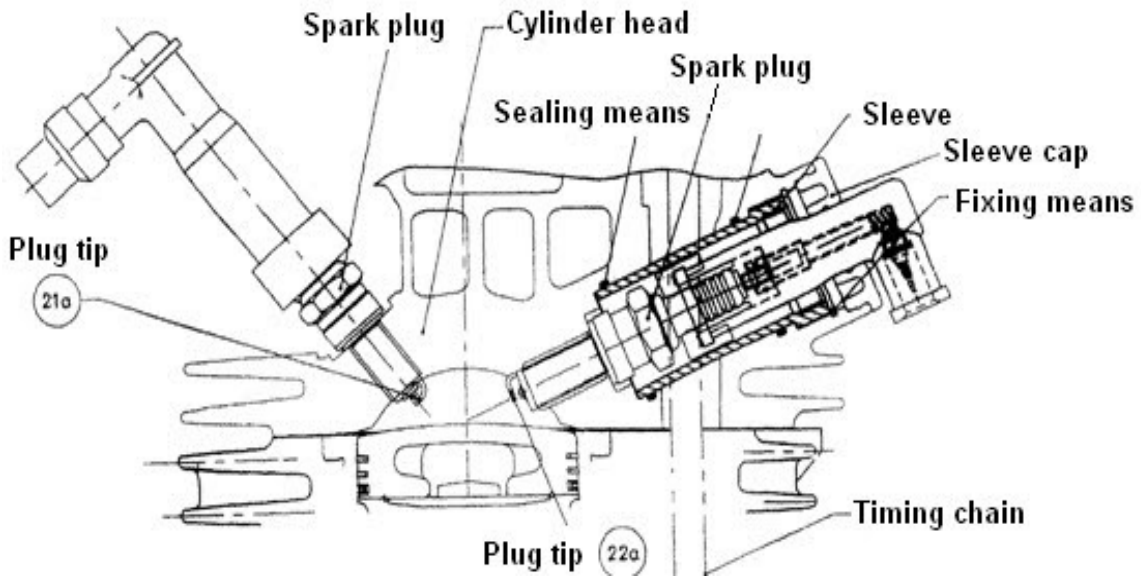
Under normal conditions it is desirable to run the engine on the maximum economy mixture, viz., around 16:1 air fuel ratio. For quick acceleration and for maximum power, rich mixture, viz., 12:1 air-fuel ratio is required

5.6 - 2 - Spark Plugs: -

A spark plug is a device for delivering electric current from an ignition system to the combustion chamber of a spark-ignition engine to ignite the compressed fuel/air mixture by an electric spark, while containing combustion pressure within the engine.

Here the only change made is that the 2-sparkplug placed at the two-opposite

end of the combustion chamber at 90 degree to each other. The distance between the spark plugs depend upon the displacement of the engine. Dual spark plug is used from 135cc engines, up to high displacement engines. Because the ignition rate is double; the power is generated product and gases expands more faster which in turns push the piston more powerfully and we get better pickup and because approximately all the fuel being ignited at once we get better fuel efficiency as well.



5.7 – 2 -Valves: -

A poppet valve is a valve typically used to control the timing and quantity of gas or vapor flow into an engine. It consists of a hole, usually round or oval, and a tapered plug, usually a disk shape on the end of a shaft also called a valve stem.

The portion of the hole where the plug meets with it is referred as the 'seat' or 'valve seat'. The shaft guides the plug portion by sliding through a valve guide. In exhaust applications a pressure differential helps to seal the valve and in intake valves a pressure differential helps open it.

6. IGNITION WITH A DIGITAL C.D.I.

A Digital CDI with an 8 bit microprocessor chip handles the spark delivery. The programmed chip’s memory contains an optimum Ignition

timing for any given engine rpm, thereby obtaining the best performance characteristics from the combustion chamber. Working together with the TRICSIII system, it delivers Optimum Ignition Timing for varying load conditions.

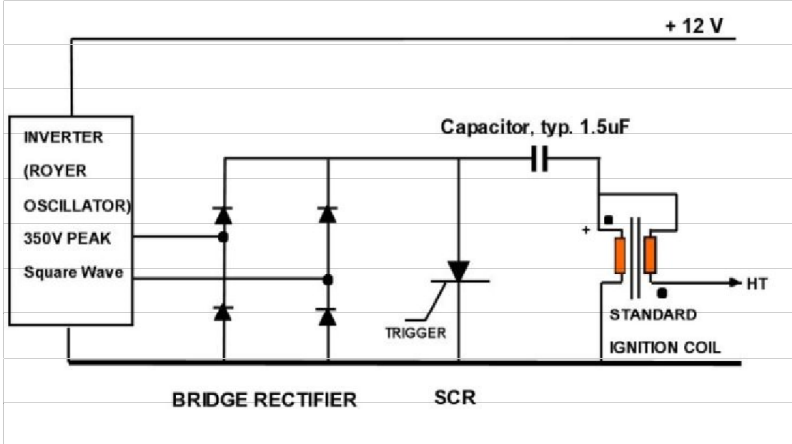
6.1 INTELLIGENT C.D.I: -

Capacitor discharge ignition (CDI) or thyristor ignition is a type of automotive electronic ignition system which is widely used in outboard motors, motorcycles, lawn mowers, chainsaws, small engines, turbine-powered aircraft, and some cars. It was originally developed to overcome the long charging times associated with high inductance coils used in inductive discharge ignition (IDI) systems, making the ignition system more suitable for high engine speeds (for small engines, racing engines and rotary engines). The capacitive-discharge ignition uses capacitor discharge current output to fire the spark plugs.

Most ignition systems used in cars are inductive discharge ignition (IDI) systems, which are solely relying on the electric inductance at the coil to produce high voltage electricity to the spark plugs as the magnetic field collapses when the current to the primary coil winding is disconnected (disruptive discharge). In a CDI system, a charging circuit charges a high voltage capacitor, and at the instant of ignition the system stops charging the capacitor, allowing the capacitor to discharge its output to the ignition coil before reaching the spark plug.

A typical CDI module consists of a small transformer, a charging circuit, a triggering circuit and a main capacitor. First, the system voltage is raised up to 250 to 600 volts by a power supply inside the CDI module. Then, the electric current flows to the charging circuit and charges the capacitor. The rectifier inside the charging circuit prevents capacitor discharge before the moment of ignition. When the triggering circuit receives triggering signals, the triggering circuit stops the operation of the charging circuit, allowing

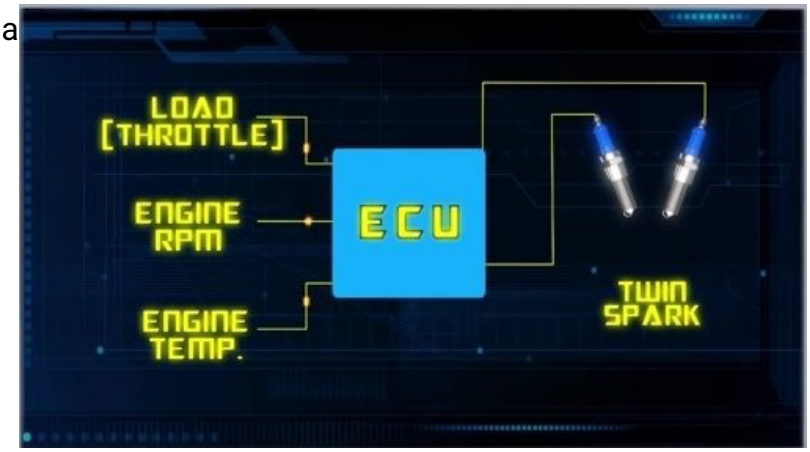
the capacitor to discharge its output rapidly to the low inductance ignition coil.



Capacitive Discharge Ignition: Functional Diagram

6.2 TRICS III:-

Throttle Response Ignition Control System III Generation. It is a means of controlling the ignition by operating the throttle. Depending on the needs of the rider whether it be cruising, acceleration or max speed, the ignition requirements constantly change. Based on a particular amount of throttle opening, the magnetic field generated by the magnet opens or closes the reed switch. The reed switch is connected to the Digital C.D.I., which signals the C.D.I. to change / switch, the desired ignition advance timing maps. This helps in achieving a good balance between drive ability and optimum ignition spark advance, resulting in an almost perfect ignition spark a



7. WORKING OF DTS-I ENGINE

The working of DTS-i engine is very similar to 4-stroke engine but here the only modification done is we are using two sparks plug at two ends of the combustion chamber. Which require less time to reach the farthest position of the combustion chamber and optimize the combustion chamber characteristics. There are some advance technology used in DTS-i engine which makes it more powerful than the conventional single sparkplug 4-stroke engine like 1.CDI technology 2. Tricks iii technology.

The cycle of operation in a four-stroke engine is completed in two revolutions of crank shaft or four strokes of piston. Stroke is defined as the distance traveled by the piston from one of the dead centers to the other dead center. It is also equal to two times the crank radius. Hence in a four-stroke engine work is obtained only during one stroke out of the four strokes of the piston required to complete one cycle.

7.1 Suction stroke: To start with the piston is at or very near T.D.C. and the inlet valve is open and exhaust valve is closed. As the piston moves from T.D.C. to B.D.C. rarefaction is formed in the cylinder which causes the charge to rush in and fill the space vacated by the piston. The charge consists of a mixture of air and petrol prepared by the carburetor. The admission of charge inside the engine cylinder continues until the inlet valve closes at B.D.C.

7.2 Compression stroke: Both the valves are closed and the piston moves from B.D.C. to T.D.C. The charge is compressed up to a compression ratio of 5:1 to 9:1 and pressure and temperature at the end of compression are about 6 to 12 bar and 250º C to 300º C respectively.

7.3 Working, Power or Expansion stroke: When the piston reaches T.D.C. position, or just at the end of compression stroke, the charge is ignited by causing an electric spark between the electrodes of two spark plug, which is

located 90 degrees each other in the walls of cylinder head. During combustion the chemical energy of fuel is released and there is rise in temperature and pressure of gases. The air-fuel mixture gets ignited in a way that creates two flame fronts and, therefore, a reduction in flame travel of the order of 40 per cent is achieved. A fast rate of combustion is achieved leading to faster rise in pressure. The obvious outcome of this is more torque, better fuel efficiency and lower emission. Now the combustion products expand and push the piston down the cylinder. The reciprocating piston motion is converted into rotary motion of crankshaft by a connecting rod and crank. During expansion the pressure drops due to increase in the volume of gases and absorption of heat by cylinder walls.

7.4 Exhaust stroke: Theoretically exhaust valve opens at the end of working stroke when the piston is at B.D.C. position. But actually, exhaust valve begins to open when about 85% of the working stroke is completed. A pressure at this instant force about complete the burnt gases into the exhaust manifold at high speed. During this stroke the pressure inside the cylinder is slightly above the atmospheric value. Some of the burnt gases are however left in the clearance space. The exhaust valve closes shortly after the piston reaches T.D.C. The inlet valve opens slightly before the end of exhaust stroke and the cycle repeats.

7.5 Detonation in DTS-I engine: In petrol engine major problem is DETONATION. Detonation occurs when the flame travel time from center to the side walls increases the self-ignition temp. time. It happens often in the engines having single spark plug.

This is overcome by the DTS-i by having 2 spark sources on the opposite sides which reduces the flame travel as stated above. Hence no detonation.

EXAMPLE: -

Pulsar DTS-i delivers out class performance in its class of vehicles. The performance characteristics are –

High Speed 180cc = 16.01 Ps (11.77kw) at

8000rpm	
	150cc = 13.02 Ps (9.57kw) at
8500rpm	
Greater torque	180cc = 14.72 N- m at 6500rpm
	150cc = 11.68 N-m at 6500rpm
Max. Speed	180cc = 127 Kmph
	150cc = 120 Kmph

8.ADVANTAGES

- Less vibrations and noise
- Long life of the engine parts such as piston rings and valve stem
- Decrease in the specific fuel consumption
- No over heating
- Increase the Thermal Efficiency of the Engine & even bear high loads on it.
- Better starting of engine even in winter season & cold climatic conditions or at very low temperatures because of increased Compression ratio.
- Because of twin Sparks the diameter of the flame increases rapidly that would result in instantaneous burning of fuels. Thus, force exerted on the piston would increase leading to better work output.

9. DISADVANTAGES

- There is high NOx emission
- If one spark plug gets damaged then we have to replace both
- The cost is relatively more

10. APPLICATIONS

It uses in automotive engines. In India Bajaj has patented for dts-i technology. At present platina, xcd125, 135, discover150, pulsar135, 150, 180, 200, 220 etc. are using the dtsi(digital twin spark ignition system). Which means the petrol enters into the cylinder burns more efficiently.

Hence the application of these technologies in the present-day automobiles will give the present generation what they want i.e. power bikes with fuel efficiency. Since these technologies also minimize the fuel consumption and harmful emission levels, they can also be considered as one of the solutions for increasing fuel costs and increasing effect of global warming.

The perfect Combustion in Internal Combustion engine is not possible. So for the instantaneous burning of fuels in I.C. engine twin spark system can be used which producing twin sparks at regular interval can help to complete the combustion.

11.CONCLUSION

In the world of new high-speed cars and bikes to achieve maximum engine power, top fuel efficiency and minimum emission levels during all type of operating condition. The digital spark ignition is the best alternative for conventional ignition control. Electronic control Unit glves accurate timing for all operating condition. At the same time use of two spark plug improves thermodynamic efficiency and power available. At the same time, it reduces the maintenance cost due to fewer moving parts in turn less friction and wear. It also good solution to reduce pollution since it minimizes emission levels. Also, it is flexible enough in mounting location. This is important because today's smaller engine compartment. Thus, it is better in all areas like power, speed, efficiency and clean emission and hence it has brought a new evolution in automobile industry.

PRINCIPAL
AMC ENGINEERING COLLEGE
BENGALURU-560 083.