Delving

Journal of Technology and Engineering Science

ISSN 09755289
December 2018
Vol. 2 Issue 1

Acropolis Institute of Technology & Research Indore
www.acropolis.in

Journal of Technology and Engineering Science

ISSN 09755289

Prof. (Dr) Kamal K Sethi
Editor in Chief
Professor & Head Information Technology
Acropolis Institute of Technology & Research Indore

Prof. Praveen Bhanodia
Editor
Associate Professor Information Technology
Acropolis Institute of Technology & Research Indore

Editorial Office
Acropolis Delving
Acropolis Institute of Technology & Research Indore
Bypass Road, Manglia Square, Manglia,
Indore, Madhya Pradesh – 453771
India

Corresponding Job on Multi-Core Processors

Anil Malviya[1]
Patel College of Science and Technology, India
malviya.anil2@gmail.com

Pritesh Jain[2 *]
Patel College of Science and Technology India
Pritesh.Arihant@Gmail.Com

Abstract

The era of personal computer arrived and more applications are generated and thus the end user requires faster and more capable system. In single core architecture, we can achieve the speedup by increasing clock speeds in single ISA (Instruction Set Architecture) but there is a limitation to increase clock frequency. Another way to achieve speedup is to add multiple processing cores to a single chip called multicore architecture. Generally there are two types of multicore processors, i.e. Symmetric Multicore Processor (SMP) and Asymmetric Multicore Processor (AMP). All the operating systems like Linux, Unix and Windows are designed keeping in view the SMP architecture but due to efficiency and less power consumption AMP processors are becoming more popular nowadays. In Amp, due to asymmetric performance of cores, task having low priority may be scheduled on high-performance core and task having low priority may be scheduled on less-performance core. So, scheduling of tasks on appropriate core is necessary. To schedule tasks in appropriate core, we will use priority class based task scheduling approach. In this approach, first we find the rank of the cores(faster to slower) and according to that task having high priority will map to faster cores and the task with low priority will map to slower cores. Map function will divide tasks into classes, based on their priority according to number of cores present in the system.

1. Introduction

The era of personal computer arrived and more applications are generated and thus the end user requires faster and more capable system. In single core architecture, we can achieve the speedup by increasing clock speeds in single ISA (Instruction Set Architecture) but there is a limitation to increase clock frequency. Another way to achieve speedup is to add multiple processing cores to a single chip called multicore architecture. A multicore processor is a single computing component with two or more independent actual processing units called cores, which are the units that execute program instructions. The multiple cores can run multiple instructions at the same time, increasing overall speed of program execution. The cores are typically integrated onto a single integrated circuit die known as chip multiprocessor or CMP.

The use of multi-core architecture has rapidly increased to develop processors as it improves speed by adding multiple processing cores to a single chip. However, adding multiple cores to a single chip processor have many challenges related to memory, cache coherence, power consumption, load imbalance among multiple cores etc? Among these challenges the assignment of task according to their priority is one of the major challenges. To schedule task according to their priority in asymmetric multi-core, the work assign on cores must be according to the priority of the task. If the task with high priority gets lesser core then the overall execution time of such tasks increases. In this time, proper ranks of core according to their efficiency should be known in advance. So, the tasks will schedule according to their priority.

2. Related Work

2.1 Task Snatching Technique

In Task Snatching Technique, the idle fast core snatches the tasks from the slow core. But in this technique is a fast core snatch the task from slow core with having less execution time and another slow core having a task with more execution time. This gives less performance compared to the optimal task scheduling [1].

2.2 CAMP

2.2.1 Utility Factor

In CAMP, a new metric Utility Factor (UF), which produces a single value that approximates how much an application, will improve its performance if its threads are allowed to occupy all the fast cores available on an AMC. The metric is designed to help the scheduler picks the best threads to run on fast cores in non-trivial cases. By comparing utility factors across threads the scheduler should be able to identify the most profitable candidates for running on fast cores. [8]

2.2.2 Scheduling Algorithm

After the utility factors of all the applications are calculated, CAMP decides which threads to place on cores of different types based on their individual utility factors. In order to achieve this purpose, threads are categorized into three classes: LOW, MEDIUM, and HIGH according to their utility factors. Threads falling in the HIGH utility class will be allocated to fast cores. As we discussed above, CAMP is a thread-level scheduler. Because tasks in the same task-based programs can often achieve similar speedup ratios on fast cores, CAMP is not applicable to improve the performance of a single parallel program. Therefore, CAMP did not considered the scheduling problem in parallel applications that this chapter will address in AMC.

2.3 Bias Scheduling

Bias scheduling which matches threads to the right type of cores through dynamically monitoring the bias of the threads in order to maximize the system throughput. In this work, each application is given a bias, which reflects the core type that best suits its resource needs. By dynamically monitoring application bias, the operating system is able to match threads to the core type that can maximize system throughput. Bias scheduling takes advantage of this by influencing the existing scheduler to select the core type that bests suits the application when performing load balancing operations [10]

2.4 Speed-Based Balancing

Speed balancing algorithm to manage the migration of threads so that each thread has a fair chance to run on the fastest core available. Instead of balancing the workloads, the algorithm balances the time of a thread executing on faster and slower cores [4]

3. Detailed Problem Statement

Suppose we have an Asymmetric Multicore Processor having one fast core and three slow cores C0, C1, C2, and C3 respectively. We have four tasks T0, T1, T2, and T3 to schedule on these cores. Suppose tasks T0, T1, T2, T3 takes S0, S1, S2, S3 times on slow cores and F0, F1, F2, F3 times on fast core, we can reasonably deduce that F0<S0 , F1<S1 , F2<S3 ,F3<S3.

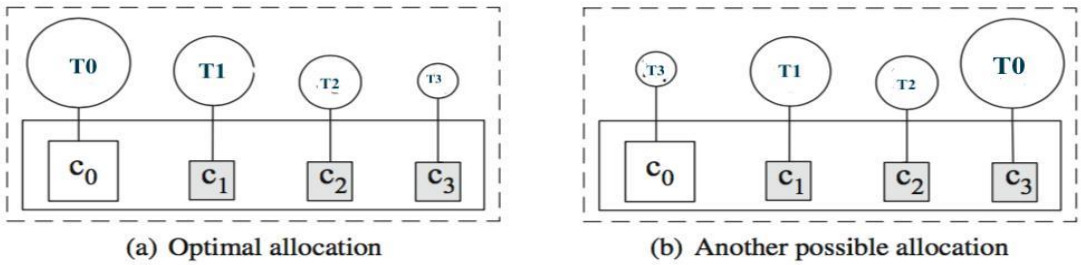

Figure-2.1: Scheduling in AMP

For easing the problem we further assumes that T0 > T1 > T2 > T3 and F0 > S1 > S2>S3. If the tasks are allocated according to optimal task scheduling means allocating the tasks such as bigger core get big task, then overall completion time (make span)

$$T_{opt} = \max(F0, S1, S2, S3) = F0$$

Because F0 < S0 we can say that Topt < S0.

In another traditional allocation technique the task T0 is scheduled on slow core C3 and task T3 is scheduled on fast core C0 . The makespan for traditional allocation is such as

$$T_{trad} = \max(S0, F1, S2, S3) \geq S0 > F0.$$

Obviously, allocating a long task to a slow core would often degrade the overall performance seriously in traditional task scheduling policies. Allocation of process to a particular core in such a way that the core having high priority will schedule on core with having high efficiency. Allocation of task in optimal way is NP hard problem, and due to change in run time behavior we can't allocate the task to appropriate core but we can schedule them in near optimal way.

4. Proposed System Architecture of Task Scheduling

The priority based task scheduling in Asymmetric Multicore Processor, in which we have multiple task (T_1,T_2,....,T_{n-1},T_n) in ready queue and having n cores (C_0 to C_n,) such that (C_0>C_1,C_1>C_2.....C_{n-1}>C_n,) and we have to map every function in such a way that the task having high priority will be scheduled on fast core and task having less priority will be scheduled on slow cores according to their priorities.

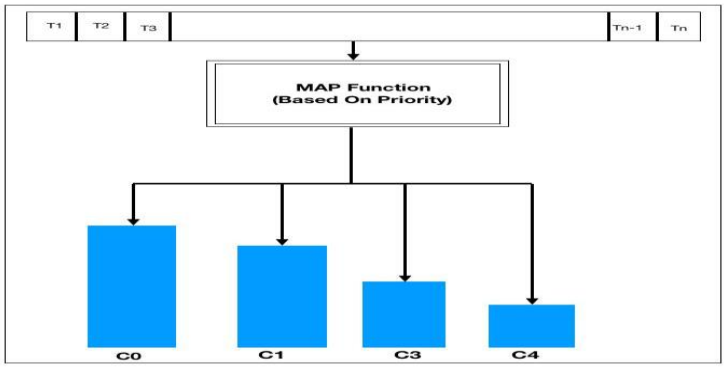

Figure-3.1: Task Scheduling

Map function is work in such a manner that it generates number of priority class equal to number of the cores. If we have three types of Cores then it maps tasks in three priority classes and according to their priority task is allocated to that core.

5. Proposed Algorithm

To schedule tasks in appropriate core, we will use. In this approach, task having high priority will map to faster cores and the task with low priority will map to slower cores. Proposed Algorithm *parallel priority class based task scheduling* approach

1. *Initiate task priorities (define no of task and priority of that task) (T1>T2>t3>T4)*

2. *Initiate CPU core load (define no of CPU core with working load) (C1>C2>C3>C4) (In order to fast to slow core)*
3. *Initiate main parallel branches (# pragma omp parallel shared ( ))*
4. *Initiate parallel section (Each task execute in separate parallel section) (# pragma omp section)*
5. *Define cpuset and current thread Cpu_set_t cpuset; pthread_t thread;*
6. *Initialisation of theread Create thread of current task Thread = pthread_self ()*
7. *Initialisation of cpuset CPU_SET(id, &cpuset);*
8. *Initialisation of priority to threads*
9. *Assign Processes to cores (Task binding to cpu core)*
   a. Set cup mask
   b. Set cpuAffinity
   (pthread_setaffinity_np (thread, sizeof(cpu_set_t), &cpuset))

6. Proposed Approach

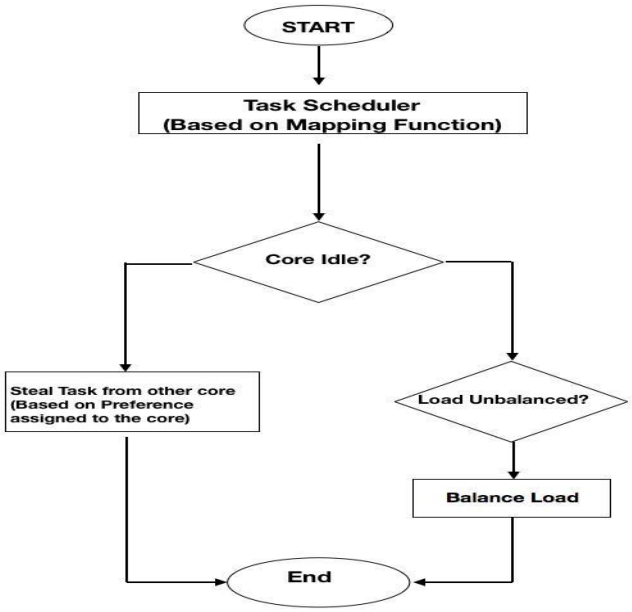

Figure-6.1 Flow Diagram of Proposed Approach

7. Result

In experiment result we saw the time difference in different case. When we assign the high priority task to the fast core performance increases.

Table 7.1 Multiple Task depend on Multiple Core bases use

	Core1	**Core2**	**Core3**	**Core4**
Task 1	13.7039	12.741		13.6703
Task 2	14.0885	14.7024	15.381	
Task 3		16.8961	16.9343	14.8626
Task 4	15.2716		16.5952	14.507

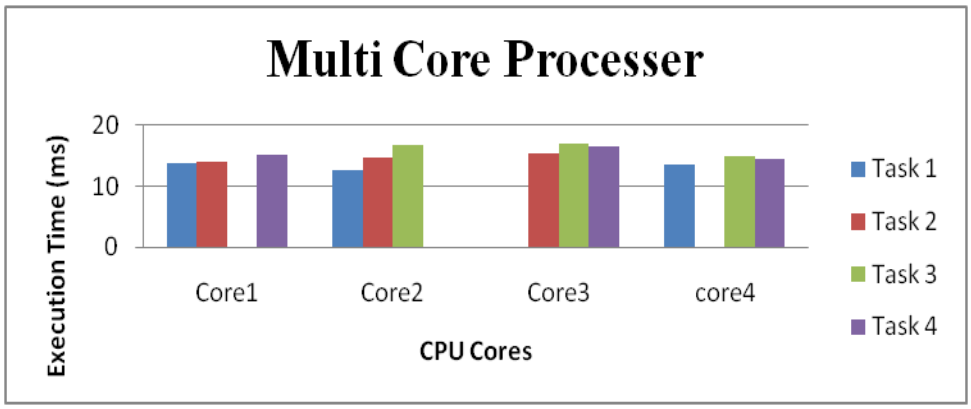

Figure-7.1: Multiple Task depend on Multiple Core bases use

8. Conclusion

Parallel programming is a difficult task. An efficient parallel implementation must manage a set of features that are not present in a similar sequential implementation. Many programming models have been proposed to tackle the complexity this introduces. These range from almost totally implicit techniques, where the compiler or runtime system makes most of the parallelization decisions, to explicitly-parallel models where the programmer has full control.

Implicit approaches remove the need to precisely specify all details of the desired parallel behavior. This is acceptable when the application is very irregular or where the programmer has little experience of parallel implementation. However for problem domains whose parallelization is well understood, a skilled practitioner can produce better implementations in a programming model which permits more low-level control. Unfortunately, explicitly-parallel programs tend to be poorly structured; the management of parallel features is tangled with computational components throughout the program. This leads to code that is difficult to understand, debug and maintain, while the presence of ma- chine specific details reduces the portability of the code.With effort, the programmer can overcome these drawbacks. A more serious problem is that explicit parallelism makes it more difficult to produce good implementations in the first place. The programming model is so cumbersome that prototyping, reasoning about correctness, and performance prediction is often seen as not being worthwhile. The paper proposes an approach to schedule tasks in appropriate core; we will use priority class based task scheduling approach. In this approach, task having high priority will map to faster cores and the task with low priority will map to slower cores. Map function will divide tasks into classes, based on their priority according to number of cores present in the system.

9. Future Work Directions

As seen from our conclusions, despite the huge research efforts already put in simplifying multi- core programming, there are still significant gaps to be filled. Therefore, our future work targets improving the framework itself, adding the required tools, and testing it on multiple applications.

References

[1] M.Benderand M.Rabin, “Scheduling Cilk multithreaded parallel programs on processors of different speeds,” in Proceedings of the 12nd annual ACM Symposium on Parallel Algorithms and Architectures., pp. 13-21 ACM, 2000.
[2] M. Bhadauria and S. McKee, “An approach to resource-aware co-scheduling for cmps,” in Proceedings of the 24th ACM International Conference on Supercomputing. ACM, 2010, pp. 189–199.
[3] C. Bienia, S. Kumar, J. P. Singh, and K. Li. The parsec benchmark suite: characterization and architectural implications. In Proceedings of the 17th international conference on Parallel architectures and compilation techniques, pages 72–81. ACM, 2008.
[4] S. Hofmeyr, C. Iancu, and F. Blagojevic, “Load balancing on speed,” in Proceedings of the 15th ACM SIG PLAN symposium on Principles and Practice Of Parallel Programming. ACM, 2010, pp. 147–158.

[5] Q. Chen, M. Guo, Q. Deng, L. Zheng, S. Guo, and Y. Shen. Hat: history-based auto-tuning mapreduce in hetero geneous environments. The Journal of Supercomputing, pages 1–17, 2013.
[6] M.De Vuyst, R.Kumar, and D.Tullsen, " Exploiting unbalanced thread scheduling for energy and performance on a cmp of smt processors," in Proceedings of the 2006 IEEE International Parallel and Distributed Processing Symposium. IEEE, 2006, pp. 10–20.
[7] Y. Guo, R. Barik, R. Raman, and V. Sarkar, "Work-first and help-first scheduling policies for async-finish task parallelism," in Proceedings of the 2009 IEEE International Parallel and Distributed Processing Symposium. IEEE Computer Society, 2009, pp. 1–12.
[8] J.C.Saez, M.Prieto, A.Fedorova, and S.Blagodurov. "A comprehensive scheduler for asymmetric multicore sys tems." In Proceedings of the 5th European conference on Computer systems, pages 139–152. ACM, 2010. M.E.(Computer Engg.) Page 28 Comp. Engg. Dept., SGSITS, Indore REFERENCES
[9] T. Fahringer, "Optimisation: Operating System Scheduling on multi-core architectures", seminar parallel compu- ting.
[10] D. Koufaty, D. Reddy, and S. Hahn, "Bias scheduling in heterogeneous multi-core architectures," in Proceed ings of the 5th European conference on Computer systems. pp.125–138 ACM, 2010.
[11] N.Lakshminarayana, J.Lee, and H.Kim, "Age based scheduling for asymmetric multiprocessors," in Proceed ings of the Conference on High Performance Computing Networking, Storage and Analysis. ACM, 2009, p. 25.
[12] M.Mahoney, "Data compression programs", 2013.
[13] http://www.kernel.com
[14] https://msdn.microsoft.com/en-us/library/windows/
[15] https://www.kernel.org/doc/readme/
[16] Kallia Chronaki, Alejandro Rico, Marc Casas, Miquel Moreto, Rosa M. Badia, Eduard Ayguade, Jesus Labarta, and Mateo Valero, "Task Scheduling Techniques for Asymmetric Multi core Systems", IEEE, 2016
[17] Robert Love,"Linux Kernel Development", third edition, Addison-Wesley publications.
[18] Stephen W. Keckler, Kunle Olukotun, H. Peter Hofstee, "Multicore Processors and Systems", Springer edition.
[19] Chengzhong Xu, Francis C. M. Lau, "Load Balancingin Parallel Comuters" demic Publishers.

Investigations for Rain Water Harvesting for Acropolis Institute of Technology and Research

Dr. Satish Kumar Sharma[1*], Priya Chauhan[1] Nupoor Kumbhkar[2] Dhruv Sharma[3]
Acropolis Institute of Technology & Research India[1, 2, 3]

Abstract

Investigation in respect of sub soil type, hydrology, topography etc was carried out at AITR, Magalia to work out suitable rain water harvesting scheme for the campus. The paper reports findings on the subject.

Keywords: Contours, Permeability, Soil Classification, Runoff, Harvesting.

1. Introduction

Scarcity of water is a global issue. India is also facing problems of scarcity of water particularly in hot weather at Indore. Water is getting depleted year after year. There is a need and necessity to avoid wastage and to conserve water. The present work is an effort in this direction for augmenting supply of water for Acropolis Institute of Technology & Research.

2. Review of Literature

Some of the important works reviewed were as under:
Pal & Sharma[1] (the first author of the current paper) reported about quantitative and qualitative improvements in ground water due to rain water harvesting in Cooch Behar, West Bengal. Kumar Dinesh et. al.[2] identified critical issues involved in rain water harvesting. They stated that economic evaluation posed several complexities. Lubina Handia et. al.[3] reported a case study of urban Zambia. They designed the system based on the mass curve analysis for storage & used a rational formula for gutters.

Che-Ani[4] discussed about implementation of rain water harvesting system in Malaysia & concluded that it would improve quality of living in Malaysia. Julius[5] reviewed the methods & design of rain water harvesting systems. He recommended implementation of rain water harvesting system on war footing. Kumar,R.[6] focussed on adopting farm ponds for harvesting of water. Amiraly,A. et. al.[7] evaluated the rain water harvesting system for the old city of Ahmedabad, Gujrat. Ganguli,R. et. al.[8] attempted to focus on effectiveness of a rain water harvesting system for Shimla region of Himachal Pradesh.

3. Methodology of Rain Water Harvesting for AITR Campus, Manglia: Various studies & investigations included the following:

1. Surveying using Total Station to develop contour maps.
2. Geotechnical investigations.
3. Hydrological survey regarding surface runoff.
4. Study of various alternate schemes in respect of roof water harvesting (figure- 1)

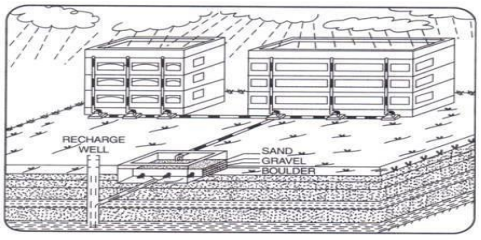

Figure-1: System of roof water harvesting

i. harvesting water in abandoned as well as existing wells (figure - 2(a & b))

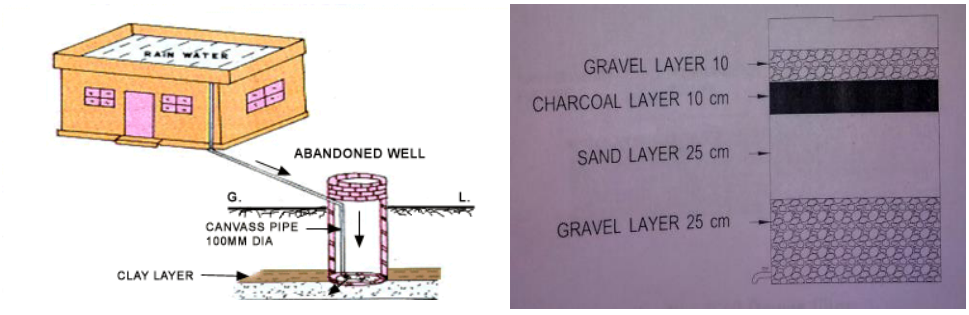

Figure-2 (a & b): System of rainwater harvesting through a well

5. Development & design of a suitable rainwater harvesting system for AITR campus keeping results of various investigations & feasibility in view.

4. Investigations carried out for Rain Water harvesting system:

1. Development of contour maps: Layout of AITR campus is shown below (Fig.1) Total Station was used for developing contour plans (Fig.2).AutoCAD was used for plotting elevations & contours. Rooftops, paved & grassy lands were surveyed. The highest elevations were identified to work out discharge. Positions of pipes were

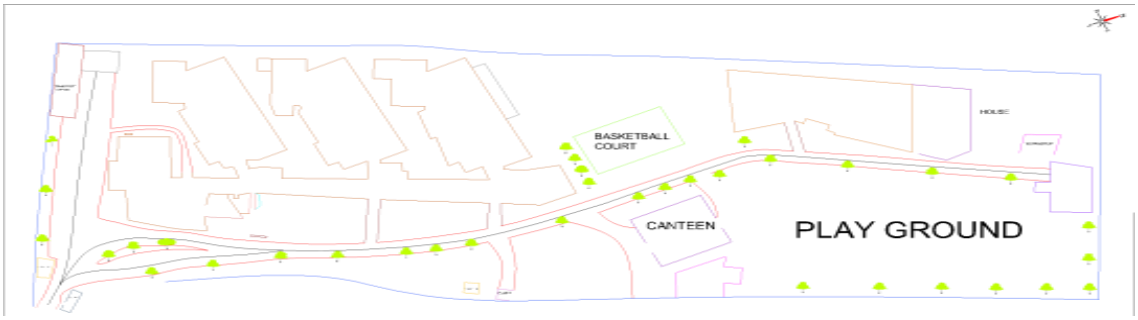

Figure-4: Layout of AITR campus identified for collection of rooftop water

Figure-5: Survey work in progress

2. Geotechnical investigations: A bore hole was drilled in the lawn behind the canteen. Disturbed as well as undisturbed samples were collected. Soil was classified as clay with high plasticity mixed with equal proportions of sand & gravel in general. Soft rock was encountered at 2.7 m depth below ground level. Water table was observed at 115m depth. A pipe 38 mm dia was inserted to conduct permeability studies. The following relationship was used for assessing permeability of soil:

$$k = \frac{3.14r}{5.5(t2 - t1)} * \log\frac{H1}{H2} - - - - - - - - - - (1)$$

Where,
r = radius of pipe.
H_1 = height of column of water from base of pipe at time t_1.
H_2 = height of column of water from base of pipe at time t_2.
Average value of coefficient of permeability was assessed to be $1.04*10^{-5}$ mm/sec.

3. Calculation of surface runoff:
Total roof area = 8831 m^2
Total paved area = 2755 m^2
Total grass land area = 28069 m^2
Sum total of all the three types of catchment areas = 38555 m^2
The total discharge,Q (m^3/sec) was assessed using the equation:

$$Q = CIA - (2)$$

Where,
A = catchment area (m^2)
I = intensity of rainfall (m/sec)
C = runoff coefficient
Q was worked out to be
0.055 m^3/sec from rooftop,
0.017 m^3/sec from paved areas,
0.037 m^3/sec from grassy lands.
Total Q = 0.11 m^3/sec

Adopting a self cleaning velocity V of 0.9 m/sec, rugosity coefficient n of 0.012 and slope, S of 1/180 & using Manning-Kutter's equation

$$Q = \frac{AR^{2/3}S^{1/2}}{n}$$

radius of semi circular pipe was worked out as 0.26 m (rounded off as 0.3 m). This size of pipe was seen to cater to the discharge (as worked out above) also.

5. Discussions & Conclusions

A suitable system of rain water harvesting is under development as per investigations briefly reported above. Rain water would be fed into abandoned as well as the working wells after suitably treating the water.

Acknowledgements

Authors are grateful to Er. Ashok Sojatia,Chairman (BOG),AITR for his guidance at various stages of this project. Thanks are expressed for Dr.S.K.Sohani, HOD, Civil Engineering, AITR for his critical comments at various stages. Some students who made substantial contribution were Aayush Singh Rathore and Adnaaan Lathsahab. This work is a part of work done by Special Interest Group created in AITR.

References:

[1] Pal, S. and Sharma, S. K., 'Rain Water Harvesting and Artificial Recharge', Journal of Indian Building Con gress, pp. 299-303, vol. 12, Nov.2005 New Delhi.

[2] Kumar Dinesh et al., 'Land Use and Water Resources' pp. 1-17, 2006.

[3] Lubina Handia et.al., 'Physics and Chemistry of the Earth' vol. 28 pp. 893-896, 2003.

[4] Che-Ani A.I. European Journal of Scientific Research, vol.34, pp.132-140,2009.

[5] Julius J.R. and Prabhavathy,publication of Department of Civil Engineering,Hindustan University, Chennai.

[6] Kumar Rohitashw, Thaman S., Agrawal G. and Poonam Sharma, 'Case Study on Modelling Rainwater System Harvesting in Ibadan, Nigeria: Application to a Residential Apartment.'American Journal of Civil Engineering and Architecture', vol. 3, no. 3, pp. 86-100, 2015.

[7] Amiraly Akil, et.al.. 'Rainwater Harvesting, Alternative to the Water supply in Indian Urban Areas: the Case of Ahmedabad in Gujarat.'(CRG-628. Working Paper. 2004).

[8] Ganguly R. et.al., 'Application of Rain Water Harvesting Scheme in Shimla Region.' publication of Department of Civil Engineering, Jaypee University of Information Technology (JUIT), . Sept. 2014, Waknaghat, District Solan, Himachal Pradesh.

Prediction of Diabetes Mellitus Using Data Mining Techniques: A Review

Varsha Rathour[1]
Patel College of Science and Technology India
varsha.rathour15@gmail.com

Pritesh Arihant[2*]
Patel College of Science and Technology India
pritesh.arihant@Gmail.Com

Abstract

Data mining techniques are used to find interesting patterns for medical diagnosis and treatment. Diabetes is a group of metabolic disease in which there are high blood sugar levels over a prolonged period. This paper concentrates on the overall literature survey related to various data mining techniques for predicting diabetes. This would help the researchers to know various data mining algorithm and method for the prediction of diabetes mellitus.

Keywords: Diabetes Mellitus, Data mining, Prediction, Decision Tree, Classification

1. Introduction

Diabetes Mellitus is a chronic disease for which there is no known cure except in very specific situations management concentrates on keeping blood sugar levels as close to normal as possible without causing hypoglycemia. This can be controlled with diet, exercise and use of appropriate medications. Diabetes Mellitus occurs throughout the world and it is more in developed countries. The increase in rates in developing countries follows the trend of urbanization and life style changes, including a "western-style" diet. This is because of less awareness. The purpose of data mining is to extract useful information from large databases or data warehouses. Data mining applications are used for commercial and scientific sides [1]. Data mining is process of selecting, exploring and modeling large amounts of data in order to discover unknown patterns or relationships which provide a clear and useful result to the data analyst [2]. KDD process may consists several steps: like data selection, data cleaning, data transformation, pattern searching i.e. data mining, finding presentation, finding interpretation and finding evaluation [3].

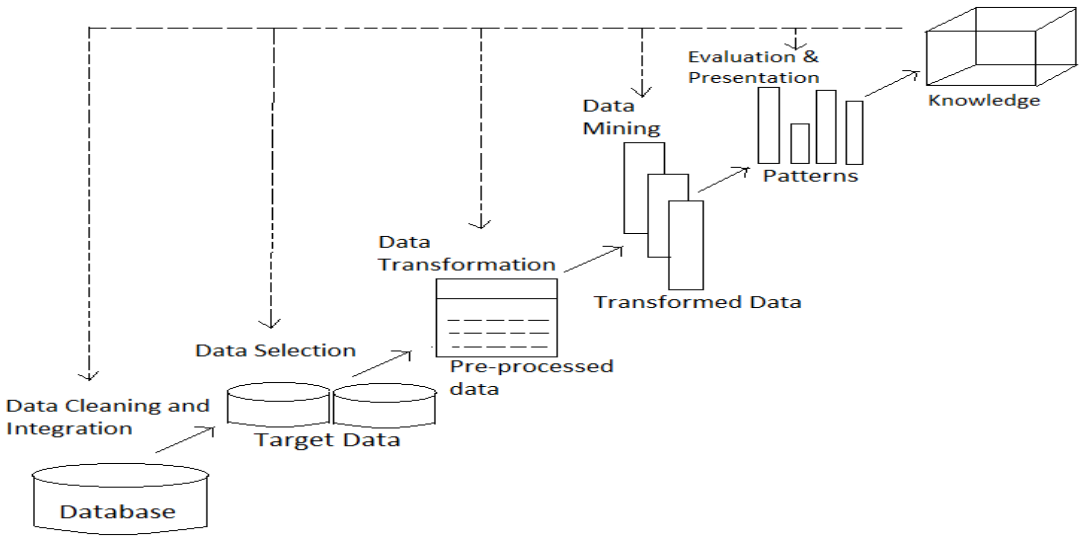

Figure 1: Knowledge Discovery Process in Data Mining

Diabetes Overview

Diabetes Mellitus (DM) is a set of related diseases in which the body cannot regulate the amount of sugar in the blood. In a healthy person, the blood glucose level is regulated by several hormones, including insulin. Insulin is produced by the pancreas, a small organ between the stomach and liver. The pancreas secretes other important enzymes that help to digest food. Insulin allows glucose to move from the blood into liver, muscle, and fat cells, where it is used for fuel.

Causes of Diabetes

Hereditary and genetics factors, Infections caused by viruses, Stress, Obesity, Increased cholesterol level, High carbohydrate diet, Nutritional deficiency, Excess intake of oil and sugar No physical exercise, Overeating, Tension and worries, High blood pressure, Insulin deficiency, Insulin resistance.

2. Literature Review

Ahmed - Using Data Mining To Develop Model For Classifying Diabetic Patient Control Level Based On Historical Medical Records. The author was motivated by the death caused by diabetes in the world which necessitated avoiding the complication of the disease. He intended to develop a new predicted model using data mining techniques which would classify diabetic patient control level based on historical medical records. The research was carried out using three data mining techniques which are Naïve Bayes, Logistic and J48. The research was implemented using WEKA application. The result showed that Logistic data mining algorithm gave a precision average of 0.73, recall of 0.744, F-measure of 0.653 and accuracy of 74.4%. Naïve Bayes gave a precision average of 0.717, recall of 0.742, F-measure of 0.653 and accuracy of 74.2%. J48 gave a precision average of 0.54, recall of 0.735, F-measure of 0.623 and accuracy of 73.5%. This proved that the logistic algorithm was more accurate than the other two. The research was limited in that only diabetes type 2 was considered. They also did not look into the discovery of appropriate features with minimal effort and validation on discovered features [4].

Ahmed - Developing a Predicted Model for Diabetes Type 2 Treatment Plans By Using Data Mining.

The author was motivated by the highly dangerous complication of chronic disease as well as the complication which required amputation of one of the parties. He intended developing a new model for classifying diabetes type 2 treatment plan which could help the control of blood glucose level of diabetic patient. He made use of J48 algorithm in conducting the experiment on 318 medical records which was collected from JABER ABN ABU ALIZ clinic center for diabetes in Sudan. The basic control information showed that 59.1% of the record was considered for Oral Hypoglycemic, 35.5% for Insulin and 5.3% for Diet. The evaluation was done using the WEKA application. The research work did not consider diabetes type 1 patients which could have been included with additional attributes. Also, the nutrition system and exercise could have been included to increase the accuracy of the system [5].

Ali et al.- Prediction of diabetes mellitus based on boosting ensemble modeling. They were motivated by the focus of aiding diabetes patients fit themselves into their normal activities of life by early predicting their state and tacking it. They intended to predict the diabetes types of patients based on physical and clinical information using boosting ensemble technique. They made use of boosting ensemble technique which internally uses random committee classifier. The architecture used was supported by integrating data management, learning, and prediction components together. The evaluation result of the technique showed accuracy gave a weighted average TP rate of 0.81, FP rate of 0.198, Precision of 0.81, Recall of 0.81, F-measure of 0.82 and ROC area of 0.82 for diabetes type 1 and 2. The research work is intended to be extended in future the integration into a cloud based clinical decision support system for chronic diseases and the inclusion of a feedback mechanism to increase the level of satisfaction of user [6].

Cole-Lewis et al.– Participatory approach to the development of a knowledge base for problem-solving in Diabetes-Self Management. The authors created a structure and component of a knowledge base in participatory design with academic diabetes educators using knowledge acquisition methods. The knowledge base validation was carried out

with the use of a scenario-based approach using inductive and deductive method. The knowledge base validation showed high level completeness and accuracy. The participatory design approached helps the capturing of implicit knowledge of practicing diabetes educator for reusability. It could enable the design of new generation of Information interventions for facilitation problem solving in diabetes self-management. The knowledge structure was not formalized as well as the relationships between its different elements. The prescription of medication was not put into consideration and the knowledge base exempted the choice of a barrier that would precede the choice of a corrective action [7].

Hempo et al.,- Personalized Care Recommendation Approach for Diabetes Patients Using Ontology and SWRL They developed the diabetes knowledge-based ontologies which was expressed in Web Ontology Language (OWL) for the description of the patient profile, the general self-care practices for diabetes patients. The ontologies were mapped and incorporated with rules which were created through the use of Semantic Web Rule Language. The semantic rule was able to enable the semantic recommendation for personalized care of patient with diabetes corresponding to each condition of the patient. Most of the system recommendation corresponding to the physician had high precision value. The system could respond very well to the needs of patient condition. The ontology web application could only be implemented by only physicians. The reasoning rules were limited to some integrated diabetes ontology [8].

Kumar and Sreejith - A Survey on Identification of Diabetes Risk Using Machine Learning Approaches. They were motivated by the machine learning approaches used in several health related studies and the fact that diabetes is a common and widely spread disease in India. They intended to survey different data mining approaches made use of in the handling of healthcare information. The carried out an exploration on the popular and effective machine learning techniques along with their advantages and disadvantages. The result showed that artificial neural network had an accuracy of 73.52%, decision tree 78.27% while regression 72.27% when used to test diabetes data. The research was limited in that only diabetes type 2 was considered. They also did not look into the discovery of appropriate features with minimal effort and validation on discovered features [9].

Mukherjee et al. - A Review of Soft computing Methods for Diabetes. They were motivated to carry out the research work based on the silently killing ability of diabetes disease which needs early prognosis of the disease so as to reduce the risk involved. This also they believe will aid the choice of diagnosis technique for prediction. They intended to compare the accuracies of diabetes diagnosis so as to find the method that produces a more efficient prediction rate of the disease. The choice of techniques made is support vector machines, decision trees, and logistic regression for the classification of pima Indian diabetes datasets. The result showed that support vector machine had an accuracy rate of 50%, decision tree 74.87% and logistic regression 77.99%. They intend in future to apply the compared techniques on attributes that were not considered for a larger number of instances [10].

Rabina and Chopra. Diabetes Prediction by Supervised and Unsupervised Learning with Feature Selection. They were motivated by the various factors which required investigation so as to diagnose diabetic patient which could make it difficult for physician. They thought it therefore to carry out a technique that was profitable for categorizing patients that are diabetic with the use of soft computing. They intended to find an approach that was better on datasets of diabetes as well as employ feature selection technique that will reduce feature and complexity of process. They carried out the research making use of WEKA application on the following techniques: Bayes Network, Naïve Bayes, Logistics, Multilayer perception, SGD and SMO. The result showed that the logistics technique had the highest accuracy with 77.7% against Bayes Network's75%, Naïve Bayes' 75.5%, Multilayer perception's 76.5%, SGD's76.7% and SMO's 76%. The result also showed that decision trees have higher potential advantage over neural networks. The research was not able to clearly denote what diabetes type was considered and neglected stating explicitly what features they were concerned with [11].

Shetty and Joshi. A Tool for Diabetes Prediction and Monitoring Using Data Mining Technique. They were motivated by the need of analysis of data with different aspects and the aggregation into information that could be useful. They intended to develop a tool that would predict and monitor diabetes with the use of data mining technique. They also intended to find out a pattern that was new and useful in the provision of information that was useful and meaningful for users who want to know their diabetes state. The research was implemented using ID3 classification algorithm which was used in identifying the disease and applied to the model for prediction. The algorithm was used in

generating decision tree from the dataset which accepted only categorical attributes for the building of the model. The evaluation result showed that the method had a 55% sensitivity, 22% specificity, 94% accuracy and error rate of 6%. The method could consider additional features for ascertaining higher level of accuracy to the model 12].

Thiyagarajan et al. - A Survey on Diabetes Mellitus Prediction Using Machine Learning Techniques. They were motivated growing interest of researchers to set up medical system which can screen great number of people for diseases that could threaten their lives. They intended to carry out a survey on machine learning techniques that have been employed in the prediction of diabetes mellitus. They also intended to propose an effective machine learning algorithm for classification so as to find the hyper-plane that was optimal which divides the various classes. They survey literatures between 2009 and 2015. Several machine learning techniques include PCGM algorithm, improved association rule mining, computational intelligent technique, preeclampsia prediction, MPSO-LSSVM algorithm and FP-growth algorithm. The investigation was done based on the performances of the techniques. They tried describing a machine learning approach to predict diabetes levels. The survey took several method of classification and ensemble them to give a new model in seeking a better result in terms of accuracy. The research work is intended to be extended for the conclusion of diabetes based on the information gathering from several locale around the world and providing more precise and general prescient model [13].

Ioannis Kavakiotisa,OlgaTsavec, Athanasios Salifoglou et all. Diabetes mellitus (DM) is defined as a group of metabolic disorders exerting significantly pressure on human health worldwide. Extensive research in all aspects of diabetes(diagnosis, etiopathophysiology, therapy, etc.) has led to the generation of huge amounts of data. The aim of the present study is to conduct a systematic review of the applications of machine learning, data mining techniques and tools in the field of diabetes research with respect to a) Prediction and Diagnosis, b) Diabetic Complications, c) Genetic Background and Environment, and Health Care and Management with the first category appearing to be the most popular. A wide range of ma-chine learning algorithms were employed. In general, 85% of those used were characterized by supervised learning approaches and 15% by unsupervised ones, and more specifically, association rules. Support vector machines (SVM) arise as the most successful and widely used algorithm. Concerning the type of data, clinical datasets were mainly used. The title applications in the selected articles project the usefulness of extracting valuable knowledge leading to new hypotheses targeting deeper understanding and further investigation in DM [14].

Manal Alghamdi, Mouaz Al-Mallah et al. The dataset contained 62 attributes classified into four categories: demographic characteristics, disease history, medication use history, and stress test vital signs. We developed an Ensembling-based predictive model using 13 attributes that were selected based on their clinical importance, Multiple Linear Regression, and Information Gain Ranking methods. The negative effect of the imbalance class of the constructed model was handled by Synthetic Minority Oversampling Technique (SMOTE). The overall performance of the predictive model classifier was improved by the Ensemble machine learning approach using the Vote method with three Decision Trees (Naïve Bayes Tree, Random Forest, and Logistic Model Tree) and achieved high accuracy of prediction (AUC = 0.92). The study shows the potential of Ensembling and SMOTE approaches for predicting incident diabetes using cardio respiratory fitness data [15].

Yukai **Li, Huling Li,and Hua Ya**o. Te focus of this study is the use of machine learning methods that combine feature selection and imbalanced process (SMOTE algorithm) to classify and predict diabetes follow-up control satisfaction data. Afer the feature selection and unbalanced process, diabetes follow-up data of the New Urban Area of Urumqi, Xinjiang, was used as input variables of support vector machine (SVM), decision tree, and integrated learning model (AdaBoost and Bagging) for modeling and prediction. Te experimental results show that AdaBoost algorithm produces better classification results. For the test set, the G-mean was 94.65%, the area under the ROC curve (AUC) was 0.9817, and the important variables in the classification process, fasting blood glucose, age, and BMI were given. Te performance of the decision tree model in the test set is relatively lower than that of the support vector machine and the ensemble learning model. Te prediction results of these classification models are sufficient. Compared with a single classifier, ensemble learning algorithms show different degrees of increase in classification accuracy. Te AdaBoost algorithm can be used for the prediction of diabetes follow-up and control satisfaction data [16].

Masatoshi Nagata, Kohichi Takai et al. This study focuses on highly accurate prediction of the onset of type-2 diabetes. We investigated whether prediction accuracy can be improved by utilizing lab test data obtained from health checkups and incorporating health claim text data such as medically diagnosed diseases with ICD10 codes and pharmacy information. In a previous study, prediction accuracy was increased slightly by adding diagnosis disease name and independent variables such as prescription medicine. Therefore, in the current study we explored more suitable models for prediction by using state-of-the-art techniques such as AdaBoost and long short-term memory (LSTM) based on recurrent neural networks. In the current study, text data was vectorized using word2vec, and the prediction model was compared with logistic regression. The results obtained confirmed that onset of type-2 diabetes can be predicted with a high degree of accuracy when the AdaBoost model is used [17].

3. Types of Diabetes

Type 1 Diabetes: It usually starts in childhood or young adulthood. The body's immune system destroys the cells that release insulin, eventually eliminating insulin production from the body. Without insulin, cells cannot absorb sugar (glucose), which they need to produce energy.

Type 2 Diabetes: It can develop at any age and usually discovered during adulthood. Now it is found that increasing number of children are being diagnosed. This can be prevented or delayed with a healthy lifestyle, including maintaining a healthy weight with regular exercise.

Gestational Diabetes: Diabetes that is triggered by pregnancy is called gestational diabetes. It is often diagnosed in middle or late pregnancy period. High blood sugar levels in a mother are circulated through the placenta to the baby and it must be controlled to protect the baby's growth and development. It creates greater risk to mother and even to the unborn baby.

4. Methodology

Publications and journals has been analyzed and data mining techniques which is given below have been applied for predicting diabetes.

Decision Tree: Decision tree is one of the popular and important classifier which is easy and simple to implement. It doesn't have domain knowledge or parameter setting. It handle huge amount of dimensional data. It is more suitable for exploratory knowledge discovery. The results attained from Decision Tree are easier to interpret and read [18].

Naive Bayes: Nave In simple terms, a naive Bayes classifier assumes that the value of a particular feature is unrelated to the presence or absence of any other feature, given the class variable. For example, a fruit may be considered to be an apple if it is red, round, and about 3" in diameter. A Naive Bayes classifier considers each of these features to contribute independently to the probability that this fruit is an apple, regardless of the presence or absence of the other features [18].

K-nearest neighbor's algorithm (k-NN): is the one of the important method for classifying objects based on closest training data in the feature space. It is simplest among all machines learning algorithm but, the accuracy of k-NN algorithm can be degraded by presence of noisy features [19].

Classification via Clustering: Clustering is the process of grouping same elements. This technique may be used as a preprocessing step before feeding the data to the classifying model. The attribute values need to be normalized before clustering to avoid high value attributes dominating the low value attributes [20].

A clinical Decision Support System based on OLAP with data mining to diagnose whether a patient can be diagnosed with diabetes with probability high, low or medium. The system is powerful because it discovers hidden patterns in the data and can, it enhances real-time indicators and discovers bottlenecks and it improves information visualization [21].

Neural Network: An artificial neural network (ANN), often just called a "Neural network" (NN), is a mathematical model or computational model based on biological neural network. Neural networks process information in a similar way the human brain does. The network is composed of a large number of highly interconnected processing elements (neurons) working in parallel to solve a specific problem [22].

In medicine, ANNs have been used to analyze blood and urine samples, track glucose levels in diabetics, determine ion levels in body fluids and detect pathological conditions [23].

Artificial Neural networks are well suited to tackle problems that people are good at solving, like prediction and pattern recognition. Neural networks have been applied within the medical domain for clinical diagnosis, image analysis and interparetation [23], signal analysis and interpretation and drug development [24].

5. Conclusion

Different approaches for the prediction of Diabetes Mellitus and its types are concentrated in this study. Data mining is a technique used to extract useful information from existing large volume of data which enable us to gain more knowledge. In this way data mining techniques are applied in health care sector in order to predict various diseases and to find out efficient ways to treat them as well.

References

[1] HianChyeKoh and Gerald Tan: Data Mining Applications in Healthcare. Journal of Healthcare Information Management, Vol 19, No 2.
[2] P. Giudici: Applied Data Mining Statistical Methods for Business and Industry. Wiley & sons, 2003.
[3] G.Piatetsky-shapiro, U.Fayyed and P.Smith: From data mining to Knowledge discovery: An overview. Advances in knowledge Discovery and Data Mining. pages 1-35, MIT Press, 1996.
[4] T. M. Ahmed, (2016a). Using Data Mining To Develop Model For Classifying Diabetic Patient Control Level Based On Historical Medical Records. Journal of Theoretical and Applied Information Technology, 87(2), 316.
[5] T. M. Ahmed, (2016b). Developing a Predicted Model for Diabetes Type 2 Treatment Plans by Using Data Min ing. Journal of Theoretical and Applied Information Technology, 90(2), 181.
[6] R. Ali, M. H. Siddiqi, M. Idris, B. H. Kang & S. Lee, (2014, December). Prediction of diabetes mellitus based on boosting ensemble modeling. In International Conference on Ubiquitous Computing and Ambient Intelligence (pp. 25-28) Springer International Publishing.
[7] H. J., Cole-Lewis, A. M., Smaldone, P. R., Davidson, , R., Kukafka, J. N., Tobin, A., Cassells, & L. Mamykina, (2016). Participatory approach to the development of a knowledge base for problem-solving in diabetes self-management. International Journal of Medical Informatics, 85(1), 96- 103.
[8] B. Hempo, N. Arch-int, , S. Arch-int, & C. Pattarapongsin, (2015). Personalized care recommendation ap-proach for diabetes patients using ontology and swrl. In Information Science and Applications (pp. 959-966). Springer Berlin Heidelberg.
[9] B.S. Kumar and Sreejith R. (2016). A Survey on Identification of Diabetes Risk Using Machine Learning Ap proaches. International Journal of innovative Research in Computer and Communication Engineering. Vol. 4, Issue 9, pp. 16752 – 16756. September 2016.
[10] S., Mukherjee, M., Thirugnanam, R., Mangayarkarasi, & T. Tamizharasi, (2015). A Review of Soft compu-ting Methods for Diabetes. International Journal, 5(2).
[11] E., Rabina, & A. Chopra, (2016). Diabetes Prediction By Supervised And Unsupervised Learning With Feature Selection.
[12] S. P., Shetty, & S. Joshi, (2016). A Tool for Diabetes Prediction and Monitoring Using Data Mining Technique. International Journal of Information Technology and Computer Science (IJITCS), 8(11), 26.
[13] C., Thiyagarajan, K. A., Kumar, & Bharathi, A. (2016). A Survey on Diabetes Mellitus Prediction Using Machine Learning Techniques. International Journal of Applied Engineering Research, 11(3), 1810-1814.
[14] Ioannis Kavakiotisa,OlgaTsavec, Athanasios Salifoglou et all. " Machine Learning and Data Mining Methods in Diabetes Research", Computational and Structural Biotechnology Journal 15 (2017) 104–116.
[15] Manal Alghamdi, Mouaz Al-Mallah , "Predicting diabetes mellitus using SMOTE and ensemble machine learn ing approach: The Henry Ford ExercIse Testing (FIT) project" , PLOS ONE |

https://doi.org/10.1371/journal.pone.0179805 July 24, 2017 pp. 1-15.
[16] Yukai Li, Huling Li,and Hua Yao , "Analysis and Study of Diabetes Follow-Up Data Using a Data-Mining-Based Approach in New Urban Area of Urumqi, Xinjiang, China, 2016-2017' , Hindawi Computational and Mathematical Methods in Medicine Volume 2018, Article ID 7207151, pages 8 ,2018
[17] Masatoshi Nagata, Kohichi Takai, Keiji Yasuda† , Panikos Heracleous, Akio Yoneyama , "Prediction Models for Risk of Type-2 Diabetes Using Health Claims" , Proceedings of the BioNLP 2018 workshop, pages 172–176 Melbourne, Australia, July 19, 2018
[18] S.Vijiyarani, S.Sudha: Disease Prediction In Data Mining Technique – A Survey. International Journal of Computer Applications & Information Technology Vol. II, Issue I, January 2013.
[19] Huy Nguyen Anh Pham and Evangelos Triantaphyllou: Prediction of Diabetes by Employing a New Data Min ing Approach Which Balances Fitting and generalization.
[20] Og uz Karan, Canan Bayraktara, Haluk Gumus_kaya, Bekir Karlıkc: Diagnosing diabetes using neural networks on small mobile devices. Expert Systems with Applications 39 (2012) 54–60.
[21] Rupa Bagdi et al : Diagnosis of Diabetes Using OLAP and Data Mining Integration. International Journal of Computer Science & Communication Networks, Vol 2(3), 314 -322.
[22] Stanfford,G.C., Kelley,P.E., Syka,J.E.P., Reynolds,W.E and Todd,J.F: Recent improvements in and analytical applications of advanced ion-trap technology. Intl. J. Mass Spectrometry Ion Processes.,1984,60: 85-98.
[23] Miller, A., Blott,B. and Hames, T: Review of neural network applications in medical imaging and signal processing. Med. Biol. Engg. Comp,1992, 30: 449-464.
[24] Weinstein, J., Kohn,K. and Grever,M.Neural: Computing in cancer drug development: Predicting mechanism of action. Science.

Implementation of Genetic Algorithm in Predicting Diabetes

Varsha Rathour[1]
Patel College of Science and Technology India
varsha.rathour15@gmail.com

Pritesh Jain[2]
Patel College of Science and Technology India
pritesh.arihant@gmail.Com

Abstract

Data Mining aims at discovering knowledge out of data and presenting it in a form that is easily compressible to humans. Data Mining represents a process developed to examine large amounts of data routinely collected. The term also refers to a collection of tools used to perform the process. One of the useful applications in the field of medicine is the incurable chronic disease diabetes. Data mining algorithm is used for testing the accuracy in predicting diabetic status. Fuzzy Systems are been used for solving a wide range of problems in different application domain Genetic Algorithm for designing. Fuzzy systems allows in introducing the learning and adaptation capabilities. Neural Networks are efficiently used for learning membership functions. Diabetes occurs throughout the world, but Type 2 is more common in the most developed countries. The greater increase in prevalence is however expected in Asia and Africa where most patients will likely be found by 2030.

Keywords: Data Mining, Diabetes, Fuzzy Systems, Genetic Algorithm (GA), Neural Networks.

1. Introduction

Health and Commonwealth Government have identified diabetes to be a significant and growing global public health problem with the expected incidence in Australia to increase from 4% to 10% by 20101. An estimated 40 million Indians suffer from diabetes, and the problem seems to be growing at an alarming rate. By 2020, the number is expected to double and reach epidemic proportions, even as half the numbers of diabetics in India remain undiagnosed. Diabetes has debilitating consequences on many of the body's vital organs if remained unchecked and controlled, the biggest problem being that of eyesight. It affects eyes, kidney, heart and every single vital organ of the body. India has the dubious distinction of being the diabetic capital of the world. Home to around 33 million people with diabetes, 19% of the world's diabetic population is from India. Nearly 12.5% of Indian's urban populations have diabetes. The number is expected to escalate to an alarming 80 million by the year 2030. Amongst the chronic diabetic complications, diabetic foot is the most devastating result. Over 50,000 leg amputations take place every year due to diabetes in India. Diabetes patients can often experience loss of sensation in their feet. Even the smallest injury can cause infection that can be various serious. 15% of patients with diabetes will develop foot ulcers due to nerve damage and reduced blood flow. Diabetes slowly steals the person's vision. It is the cause for common blindness and cataracts. Cardiovascular diseases are rising. Nearly 3.8 crore cases were detected in 2005 and experts believe the number will go up to 6.4 crore by 2015.

Fuzzy Systems is used for solving a wide range of problems in different application domains. The use of Genetic Algorithms for designing Fuzzy Systems allows us to introduce the learning and adaptation capabilities. The topic has attracted considerable attention in the Computation Intelligence community. The paper briefly reviews the classical models and the most recent trends for Genetic Fuzzy Systems. Accurate and reliable decision making in oncological prognosis can help in the planning of suitable surgery and therapy, and generally, improve patient management through the different stages of the disease. To indicate that the reliable prognostic marker model than the statistical and artificial neural network- based methods.

Genetic Algorithms (GAs) are considered as a global search approach for optimization problems. Through the proper evaluation strategy, the best "chromosome" can be found from the numerous genetic combinations. Although the GA operations do provide the opportunity to find the optimum solution, they may fail in some cases, especially when the length of a chromosome is very long. In this paper, a data mining-based GA is presented to efficiently improve the Traditional GA (TGA). By analyzing support and confidence parameters, the important genes, called

DNA, can be obtained. By adopting DNA extraction, it is possible that TGA will avoid stranding on a local optimum solution. Furthermore, the new GA operation, DNA implantation, was developed for providing potentially high quality genetic combinations to improve the performance of TGA. Experimental results in the area of digital water marking show that our data mining based GA successfully reduces the number of evolutionary iterations needed to find a solution.

Real-life data mining applications are interesting because they often present a different set of problems for data miners. One such real-life application that we have done is on the diabetic patients databases. In this paper, knowledge discovery on this diabetic patient database is discussed. A semi-automatic means for cleaning the diabetic patient database, and present a step-by-step approach to help the health doctors explore their data and to understand the discovered rules better. Generally in Asia about 47 percent of the population is diabetic. This disease has many side effects such as higher risk of eye disease, higher risk of kidney failure, and other complications. However, early detection of the disease and proper care management can make a difference. To combat this disease a regular screening program for the diabetic patients. Patient information, clinical symptoms, eye-disease diagnosis and treatments are captured into a database. This leads naturally to the application of knowledge discovery and data mining techniques to discover interesting patterns that exist in the data. The objective is to find rules that can be used by the medical doctors to improve their daily tasks, that is, to understand more about the diabetic disease.

2. Fuzzy Systems

Application of fuzzy sets theory was recognized in the field of medicine, the uncertainty found in the process of diagnosis of disease that has most frequently been the focus of applications of fuzzy set theory. The desire to better understand and teach this difficult and important process of medical diagnosis has prompted attempts to model it with the use of fuzzy sets, These models vary in the degree to which they attempt to' deal with different complicating aspects of medical diagnosis such as the relative importance of symptoms, the varied symptom patterns of different disease stages, relations between diseases themselves, and the stages of hypothesis formation, preliminary diagnosis, and final diagnosis within the diagnostic process itself. These models also form the basis for computerized medical expert systems, which are usually designed to aid the physician in the diagnosis of some specified category of diseases.

3. Genetic Algorithms

Genetic algorithm (GA) refers to a model introduced and investigated by John Holland in 1975 for adaptation processes of nature. Generally stated, a GA is any population based model that uses selection and recombination operators to generate new sample points in a search space. GA computationally utilizes a natural evolutionary process similar to the process first described by Charles Darwin in his "The Origin of Species", to solve a given problem. GA is a global search procedure that searches from one population of points to another. GA is a probabilistic search procedure, which is being frequently applied to difficult optimization and learning problems. There are two versions of the GA, namely the natural GA and the computational GA.

Genetic algorithms were inspired by the processes observed in natural evolution. They attempt to mimic these processes and utilize them for solving a wide range of optimization problems. In general, genetic algorithms perform directed random searches through a given set of alternatives with respect to the given criteria of goodness. These criteria are required to be expressed in terms of an objective function, which is usually referred to as a fitness function. Genetic algorithms require that the set of alternatives to be searched through be finite. If we want to apply Them to an optimization problem where this requirement is not satisfied, the set involved and select an appropriate finite subset. It is further required that the alternatives be coded in strings of some specific finite length which consist of symbols from some finite alphabet. These strings are called chromosomes, the symbols that form them are called genes, and their set is called a gene pool. Genetic algorithms search for the best alternative in the sense of a given fitness function through chromosomes evolution. Basic steps in genetic algorithms are shown in figure 1

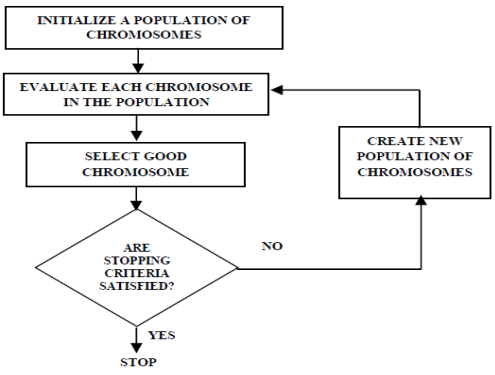

Figure-1: Step of Genetic Algorithms

Genetic Algorithms search for the best alternative (in the sense of a given fitness function) through chromosomes' evolution. Basic steps in genetic algorithms figure. First, an initial population of chromosomes is randomly selected. Then each of the chromosomes in the population is evaluated in terms of its fitness (expressed by the fitness function). Next, a new population of chromosomes is selected from the given population by giving a greater change To select chromosomes with the high fitness. This is called natural selection. The new population may contain duplicates. If given stopping criteria (e.g., no change in the old and new population, specified computing time, etc.,) are not met, some specific, genetic – like operations are performed on chromosomes of the new population. These operations produce new chromosomes, called offspring's. The same steps of this process, evaluation and natural selection, are then applied to chromosomes of the resulting population. The whole process is repeated until given stopping criteria are met. The solution is expressed by the best chromosome in the final population.

There are many variations on these basic ideas of genetic algorithms. To describe a particular type a genetic algorithm is greater detail, let G denote the gene pool, and let n denote the length of strings of genes that form chromosome. That is, chromosomes are n - tuples in Gn. The size of the population of chromosomes is usually kept constant during the execution of genetic algorithm. That is, when new members are added to the population, the Corresponding the number of old members are excluded. Let m denote this constant population size. Since each population may contain duplicates of chromosomes, we express populations by m-tuples whose elements are n-tuples from the set Gn. Finally, let f denote the fitness function employed in the algorithm.

3.1 Genetic Algorithm which is Iterative, Consists of the Following Six Steps

1. Select an initial population, $P^{(k)}$, of a given size m, where k=1. This selection is made randomly from the set G^n. The choice of value m is important. If it is too large, the algorithm does not differ much from an exhaustive search; it is too small, the algorithm may not reach the optimal solution.

2. Evaluate each chromosome in population $P^{(k)}$ in terms if its fitness. This is done by determining for each chromosome x in the population the value of the fitness function, f(x).

3. Generate a new population $Pn^{(k)}$, from the given population $P^{(k)}$ by some procedure of natural selection. We describe only one possible procedure of natural selection, which is referred to as deterministic sampling. According to this procedure, we calculate the value e(x) = mg(x) for each x in $P^{(k)}$, where g(x) is a relative fitness defined by the formula.

$$G(k) = \frac{f(X)}{\sum_{n \in p^k} f(X)} \quad \text{– – – – – – – – – – – – (1)}$$

Then the number of copies of each chromosome x in $P^{(k),}$ that is chosen for $Pn^{(k),}$ is given by the integer part of e(x). If the total number of chromosomes chosen in this way is smaller than m (the usual case), then we select the remaining chromosomes for $Pn^{(k)}$ by the fractional parts of e(x), from the highest values down. In general, the purpose of this procedure is to eliminate chromosomes with low fitness and duplicate those with high fitness.

4. If stopping criteria are not met, go to step 5, otherwise stop.

5. Produce a population of new chromosomes $P^{(k+1)}$), by operating on chromosomes in population $Pn^{(k).}$ Operations that are involved in this step attempt to mimic genetic operations observed in biological systems. They include some or all the following four operations:

A. Simple Crossover

Given two chromosomes
$X= (X_1, X_2 \ldots N_n)$, $Y=(Y_1,Y_2,\ldots\ldots\ldots\ldots\ldots,Y_n)$
And an integer $i \varepsilon N_{n-1}$ which is called a crossover position, the operation of simple crossover applied to x and y replaces these chromosomes with their offspring's,
$X' = (X1 , \ldots\ldots. X_i , Y_{i+1} , \ldots., Y_n)$,
$Y' = (Y_1, \ldots., Y_i , X_{i+1} , \ldots . , X_n)$
Chromosomes x and y, to which this operation is applied, are called mates.

B. Double Crossover

Given the same chromosomes mates x,y as in the simple crossover and two crossover positions i, $j \varepsilon N_{n-1}$ (i<j) , the operation of double crossover applied to x and y replaces these chromosomes with their offspring's,
$X' = (X_1 , \ldots\ldots. X_i , Y_{i+1} , \ldots., Y_j , X_{j+1 , \ldots\ldots , } X_n)$,
$Y' = (Y_1 , \ldots . , Y_i , X_{i+1} , \ldots . , X_j , Y_{j+1} , \ldots., Y_n)$

C. Mutation

Given a chromosome $X= (X_1 , X_2 , \ldots\ldots\ldots\ldots , X_n)$ and an integer $i \varepsilon N_n$, which is called a mutation position, the operation of mutation replaces x with
$X' = (X1 , \ldots\ldots. X_{i-1} , Z , X_{j+1 , \ldots\ldots , } X_n)$,
Where z is a randomly chosen gene from the gene pool G.

D. Inversion

Given a chromosome $X= (X_1 , X_2 , \ldots\ldots\ldots\ldots , X_n)$ and two integers i, $j \varepsilon N_{n-1}$ (i<j) , which are called inversion positions, the operation of inversion replaces x with
$X' = (X1, \ldots\ldots. X_i , X_j , X_{j-1} , X_{i+1} , X_{j+1 , \ldots\ldots , } X_n)$
E. Replace population $P_n^{(k)}$ with $P^{(k+1)}$ produced in Step 4, increase k by one, and go to Step 2.

3.2 Sample Problem

(a) k = 1 : Step 2 and 3

Chromosome in $P^{(1)}$	Integers	Fitness	g(x)	4g(x)	Number of Selected Copies
00010	2	3.75	0.068	0.272	0
01001	9	12.94	0.292	1.168	1
10011	19	15.44	0.350	1.400	2
11000	24	12.00	0.291	1.164	1

(b) k = 1 : Step 5

Chromosome in $Pn^{(1)}$	Mate (randomly Selected)	Crossover Site (randomly Select-ed)	Resulting Chromosomes in $P(^2)$
01001	10011	3	01011
10011	01001	3	10001
10011	11000	1	11000
11000	10011	1	10011

Similarly for the values of k 2, 3, values are set to calculate the fitness value, mate and the crossover site.

(c) k = 2 : Step 2 and 3

Chromosome in $P^{(2)}$	Integers	Fitness	g(x)	4g(x)	Number of Selected Copies
01011	11	14.44	0.250	0.100	0
10001	17	15.94	0.276	1.104	2
11000	24	12.00	0.207	0.828	1
10011	19	15.44	0.267	1.068	1

(d) k = 2 : Step 5

Chromosome in $Pn^{(2)}$	Mate (randomly Selected)	Crossover Site (randomly Select-ed)	Resulting Chromosomes in $P(^3)$
10001	3	2	10000
10001	4	3	10011
11000	1	2	11001
10011	2	3	10001

(e) k = 3: Step 2 and 3

Chromosome in $P^{(2)}$	Integers	Fitness	g(x)	4g(x)	Number of Selected Copies
10000	16	16.00	0.274	1.096	1
10011	19	15.44	0.265	1.060	1
11001	25	10.94	0.188	0.752	1
10001	17	15.94	0.273	1.092	1

A crossover operation is employed in virtually all types of genetic algorithms, but the operations of mutation and inversion are sometimes omitted. Their role is to produce new chromosomes not on the basis of the fitness function, but for the purpose of avoiding a local minimum. This role is similar to the role of a disturbance employed in neural networks. If these operations are employed they are usually chosen with small probabilities. The mates in the cross-over operations and the crossover positions in the algorithm are selected randomly. When the algorithm terminates, the chromosome in $P^{(k)}$ with the highest fitness represents the solution.

3.3 Natural Genetic Algorithm

The natural genetic algorithm is as follows:

- Randomly Generate an initial population M(O)
- Loop

a. Compute and save the fitness u(m) for each individual m in current population M(t).
b. Define the selection probabilities p(m) for each individual m in M(t) (so that p(m) is proportional to u(m).

c. Generate M (k+1) by probabilistically selecting individuals from M(t) to produce a new population via genetic operators.
Fuzzy Genetic Algorithm can be implemented to check the patients affected by diabetes based upon the fitness value and the accuracy chromosome value.

3.4 Diabetes

Most of the food we eat is converted to glucose, or sugar which is used for energy. The pancreas secretes insulin which carries glucose into the cells of our bodies, which in turn produces energy for the perfect functioning of the body. When you have diabetes, your body either doesn't make enough insulin or cannot use its own insulin as well as it should. This causes sugar to build up in your blood leading to complications like heart disease, stroke, and neuropathy, poor circulation leading to loss of limbs, blindness, kidney failure, nerve damage, and death.

3.4.1 General Symptoms of Diabetes

- Increased thirst
- Increased urination - Weight loss
- Increased appetite - Fatigue
- Nausea and/or vomiting - Blurred vision
- Slow-healing infections - Impotence in men

3.4.2 Types of Diabetes

Type 1 - Diabetes also called as Insulin Dependent Diabetes Mellitus **(IDDM)**, or Juvenile Onset Diabetes Mellitus is commonly seen in children and young adults however, older patients do present with this form of diabetes on occasion. In type 1 diabetes, the pancreas undergoes an autoimmune attack by the body itself therefore; pancreas does not produce the hormone insulin. The body does not properly metabolize food resulting in high blood sugar (glucose) and the patient must rely on insulin shots. Type I disorder appears in people younger than 35, usually from the ages 10 to 16.

Type II - Diabetes is also called as Non-Insulin Dependent Diabetes Mellitus (NIDDM), or Adult Onset Diabetes Mellitus. Patients produce adequate insulin but the body cannot make use of it as there is a lack of sensitivity to insulin by the cells of the body. Type II disorder occurs mostly after the 40.

Gestational Diabetes-

Diabetes can occur temporarily during Pregnancy called as Gestational Diabetes which is due to the hormonal changes and usually begins in the fifth or sixth month of pregnancy (between the 24th and 28th weeks). Gestational diabetes usually resolves once the baby is born. However, 25-50% of women with gestational diabetes will eventually develop diabetes later in life, especially in those who require insulin during pregnancy and those who are overweight after their delivery.

3.4.3 Diagnostic Tests

- Urine Test
- Fasting Blood Glucose Level
- Post Prandial Blood Sugar
- Random Blood Glucose Level
- Oral Glucose Tolerance Test
- Glycosylated Hemoglobin (HbAlc)

4. Computation Genetic Algorithm

4.1 Pseudo code of Genetic Algorithms

```
function sga ()
{
Initialize population;
Calculate fitness function;
While(fitness value != termination criteria)
{
Selection;
Crossover;
Mutation;
Calculate fitness function;
}
}
```

4.2 Generic Genetic Algorithm

Procedure GA_IPD_Run
Initialize_Population (P_{old})
// fills the chromosome of population P_{old} with 0's and 1's randomly. While termination criteria not satisfied do for each chromosome c_i in P_{old} do
Evaluate (c_i, P_{old}) // runs chromosome ci against every member of Pold includes itself to compute fitness end
Generate_New_Population (P_{new}, P_{old})
// generate new population using
P_{old} P_{old} ⟶ P_{new} end ,
End

4.3 Algorithm for Generating New Population

Procedure Generate _ New _ Population (P_{old}, P_{New})
P_{New} ⟶ 0
while Size (P_{New}) < Size (P_{old}) do
// Selection
c1 ⟵ Select (P_{old}) c2 ⟵ Select (Pold)
// Crossover
if Pc < r(.) then // return random nos. in the interval (0,1)
// Pc : Crossover Probability Crossover (c1, c2) //
Implements uniform crossover end
// Mutation
for i = 1 to chromosome_ length do
if r(.) < Pm then // Pm Mutation Probability
// Chromosome swapping each bit at the
Corresponding position with fixed probability usually 0.5 percent
c1i⟵c1i // ith bit of the Ist chromosome
end

if r(.) < P_m then c_{2i} ⟵c_{2i} end, end
P_{New} P_{New} ∇ c_1 ∇ c_2 // ∇ : Inserts the chromosome on the right hand side to the population to the left hand side. ⟶

5. Experimental Results

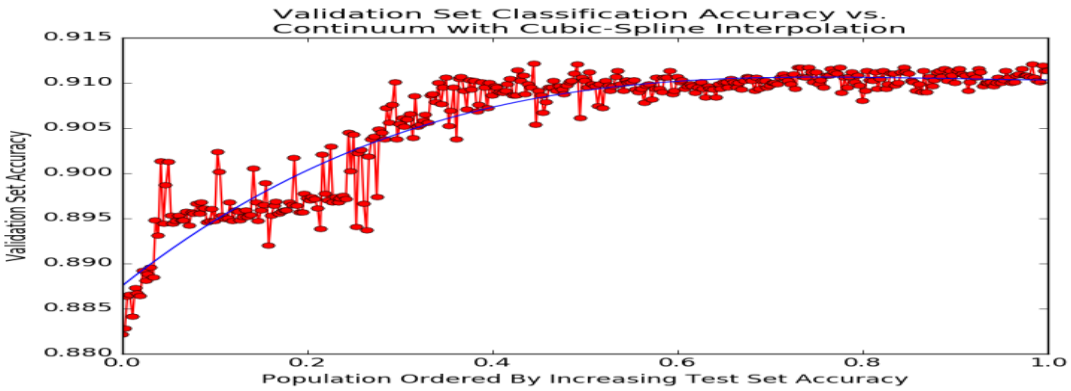

Figure 1: Validation set classification accuracy vs continuum with cubic-spline interpolation

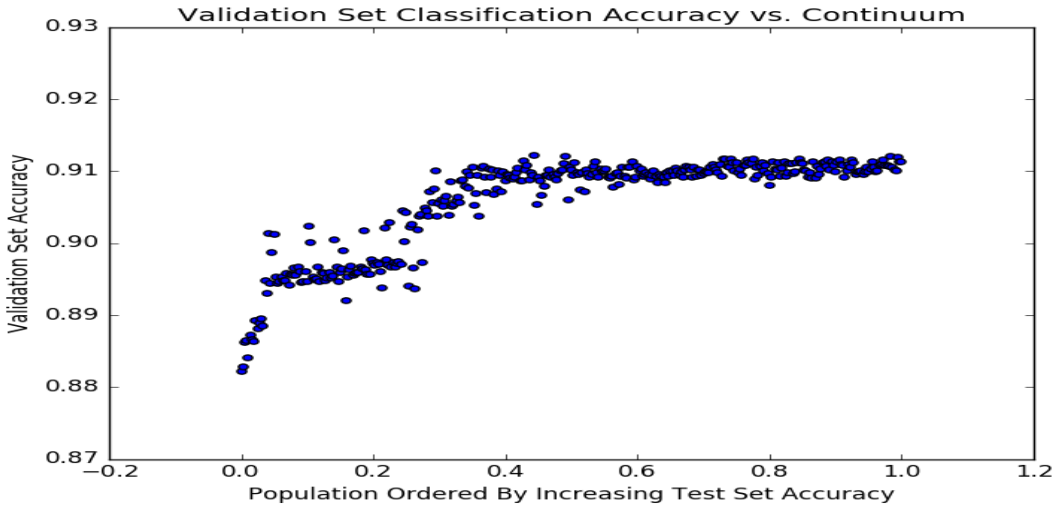

Figure 2: Validation set classification accuracy vs continuum

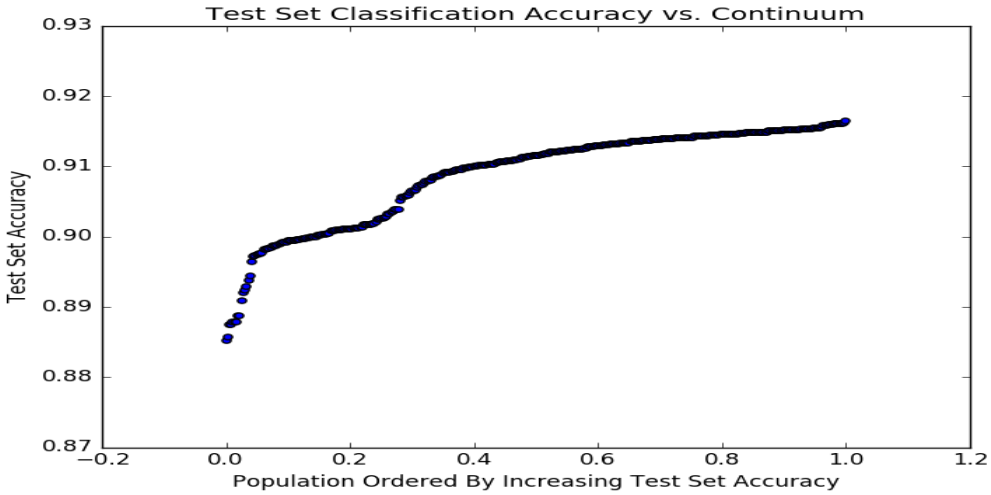

Figure 3: Classification Accuracy vs. Continuum

6. Conclusion

In order to obtain the accuracy of chromosome and to evaluate the diabetes in diabetic patient GA is implemented. The connection between fuzzy systems and genetic algorithms is bidirectional. In one direction, genetic algorithms are utilized to deal with various optimization problems involving fuzzy systems. One important problem for which genetic algorithms have proven very useful is the problem of optimizing fuzzy inference rules in fuzzy controllers. In the other direction classical genetic algorithms can be fuzzified. The resulting fuzzy genetic algorithms tend to be more efficient and more suitable for some applications. Research on complex diseases only seems to be approaching the final goal, the prevention and cure of the diseases, very slowly. Diabetes is a disease in which the body does not produce or properly use insulin. Insulin is a hormone that is needed to convert sugar, starches and other food into energy needed for daily life. The cause of diabetes continuous to be ambiguous although both genetics and environmental factors such as obesity and lack of exercise. Symptoms of low blood sugar, side effects, science of complication are to be noted else it leads to severe problems. Using GA optimization of chromosome is obtained and based on the rate of old population diabetes can be restricted in new population to get chromosomal accuracy.

References

[1] N. Nordisk, “Diabetes and you your guide to living well with diabete”, Lead Group
[2] G. Eliot , and V. Patel “How to cut out all Diabetic Problems by 50% - The Alphabet way”, Department of Di -abetes and Endocrinology Hospital, UK.
[3] Genetic mapping of complex traits: the case of Type 1 diabetes, Päivi Onkamo, Diabetes and Genetic Epidemiology Unit, Department of Epidemiology and Health Promotion, National Public Health Institute and Division of Biometry, Rolf evanlinna Institute and Finnish Genome Center Faculty of Science University of Helsinki Aca demic, 2002.
[4] Genetic Fuzzy Systems: Status, Critical Considerations and Future Directions, Francisco Herrera, Interna-tional Journal of Computational Intelligence Research. ISSN 0973-1873 Vol.1, No.1 (2005), pp.59-67.
[5] A Fuzzy Logic Based-Method for Prognostic Decision Making in Breast and Prostate Cancers, Huseyin Sek-er, Student Member, IEEE, Michael O. Odetayo, Dobrila Petrovic, and Raouf N. G. Naguib, TRANSACTIONS on Information Technology in Biomedicine, Vol. 7, No. 2, June 2003.
[7] Yi-Ta Wu, Yoo Jung An, James Gellar and Yightyng Wu, “A Data Mining Based Genetic Algorithm”, 2006 IEEE.
[8] Joseph L. Breault, MD, MPH, MS, “Data Mining Diabetic Databases: Are Rough Sets a Useful Addition?” De partment of Health Systems Management, Tulane University, Department of Family Practice, Alton Och-sner Medical Foundation, joebreault@tulanealumni.net.
[9] Y.Min, Huang, S. Chen, Cheng, “Parallel Medical Image Analysis for Diabetic Diagnosis”, Int. J. Computer Appli-cations in Technology, Vol. 22, No. 1, 2005.
[10] J. Yang and V. Honavar, “Feature Subset Selection Using a Genetic Algorithm” IEEE Intelligent Systems Ex -ploration Mining in Diabetic Patients Databases : Findings and Conclusions,

A Review on Visual Saliency Detection Based On Multi Scale K-NN Features

Gaurav Mandloi[1 *]
AITR India
gauravmandloi@acropolis.in

Pawan K Gupta[2]
AITR India
pawangupta@acropolis.in

Dr. Dinesh Jain[3]
AITR India
dineshjain@acropolis.in

Abstract

Image Retrieval is the process of getting knowledgeable information about the images that can be used in various applications based on matches and differences in face recognition, pattern recognition, and feature extractions. Now a days so many feature extraction techniques are available but so many flows and accuracy could not meet as we could find in the literature survey. We are proposing a system based on visual saliency which can extract and store the features from a given set of images and finally this database uses as input to k-NN algorithm to produce more accurate and detailed result of user query.

Keywords: Digital image, image retrieval, feature extraction, k-Nearest Neighbor algorithm (KNN), tags, patterns

1. Introduction

In content based image retrieval techniques the term “Content-based" means the search made by analyzing the contents of stored images in database rather than their associated text i.e. keywords, or other kind of descriptors. In this context the content may refers to colors, shapes, and textures features. Additionally, that can also include any other information that derived from image can be used as content of the image. The key reason behind development of CBIR is the limitation of text based image retrieval systems. Textual information about images can be easily searched using various existing technology, but this requires huge human efforts to describe image and their contents. Additionally due to lingual issues such as use of different words in different regions for description can also affect the performance of image search. Therefore search by text in image context is not much suitable.

In machine learning procedure data is perceived utilizing their compelling patterns and extricated utilizing the closeness between these patterns. To discover the conspicuous patterns among the data needed to lessen the measure of data and concentrate the genuine relationship or difference between two data instances. These relationships or differences are figured utilizing the substance of the data. Accordingly, that is a complex domain; where uncertainty and random nature of the data can be misinform the real decision or recognition design.

The exhibited work in this paper is an assessment of techniques by which the ideal properties between the data can be assessed to discover and from the ideal properties by which the nature of data and the pattern of the data can be recognized. The introduced work is an assessment of the image data and discovering the most suitable feature extraction strategy, with a specific end goal to use them in different applications.

For a legitimate understanding of the connection between the data processing and image processing, first we take a sample, assume we have a set of random documents, for ordering or fitting game plan of these documents as indicated by their domain, needed to discover some knowledge about the document substance, consequently initially needed to peruse a document and afterward assess the domains and in the given document. In the same route for discovering the suitable patterns over the given data, pre-processing, data model construction and implementation in problem is needed.

Image is a different sort of data which includes a tremendous measure of information, such as colour information, objects, edges, pixel definition, dimensions and others. Along these lines the treatment of image data is a sensitive concern to preserve the complete information. This paper addresses the various key features and properties of image

data by which the information from the image is separated and used for different applications of face recognition, image retrieval and others.

1.1 Image Retrieval

An image retrieval scheme can be described as penetrating, browsing, and recovering images from huge databases consisting of digital images. Though Conservative and communal methods of recovering images make use of adding metadata namely captioning the keywords so as to perform annotation of words. However image search can be describe by devoted technique of search which is mostly used to find images. For search images user provide the query image and the method profits the image related to that of query image [2]. Image Recovery has been provided by almost all main search engines, containing Google, Yahoo!, Bing, etc. A big number of image exploration engines mostly employ the adjacent texts nearby the images and the image names to directory the images. Because there are only two main places where anyone can enter text first in title (Name of image) and second in the tags which are proposed and implemented using web 2.0 concepts. Most of the time user fires query in the text format for search the contents over any search engine.

An image retrieval system can be defined as searching, browsing, and retrieving images from massive databases consisting of digital images. Although conformist and common techniques of retrieving images make use of adding metadata namely captioning keywords so as to perform explanation of words. However image search can be described by the dedicated technique of search which is mostly used to find images.

With the expansion of the Internet, and the accessibility of image capturing devices such as digital camera, vast amounts of images are being produced every day in dissimilar areas including distant sensing, fashion, fault prevention, publishing, medicine, construction, etc. For this purpose, the necessity for the development of capable and effective methodologies to organize large image databases for retrieval is vital so many general- purpose image retrieval systems have been developed. There are three techniques for image retrieval: text-based method, content-based method and hybrid method. This section explains in detail each method. The image retrieval system can be classified as:

- Text based Image retrieval system
- Content Based Image retrieval system

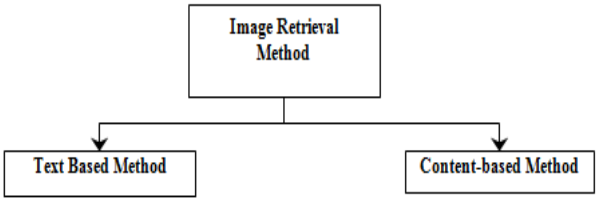

Figure-1: Taxonomy

Text Based Image Retrieval (TBIR) is presently used in almost all general-purpose web image retrieval systems today. This approach uses the text connected with an image to determine what the image contains. This text are the text adjacent the image, the image's filename, a hyperlink essential to the image, an clarification of the image, or any other part of text that can be connected with the image. Google, Yahoo Image investigates engines are instances of systems using this approach. Even though these search engines are quick and dynamic, they sometimes fail to retrieve related images; this is because of many reasons

- Firstly, there are so many unsuitable words in the nearby textual definitions, which cost in low image search accuracy rate.

- Secondly, the close text does not seem to fully clarify the semantic content of network images, which outcome in lower image search recall rate.

- The third trouble is a polysemy problem (same word can be used to submit to more than one object). Due to the inquiry polysemy, the results searcher will be unsuccessful to find images tagged with Chinese, and a Dutch searcher will be unsuccessful to find images tagged in English. This funds the inquiry must match the language of the text linked with the images.Content Based Image Retrieval is a set of techniques for retrieving semantically related images from an image database based on automatically-derived image features [3]. This target for avoiding the use of textual explanations and in its place retrieves images based on their visual relationship to a user-supplied query image or user-specified image features.

The major objective of CBIR is effectiveness during image indexing and retrieval, thereby falling the require for human involvement in the indexing method [4]. The computer must be able to retrieve images from a database without any human assumption on specific domains (such as texture vs. non texture). The major jobs for CBIR systems is similarity comparison, extracting quality of every image based on its pixel standards and defining rules for comparing images. These features become the image representation for measuring similarity with other images in the database.

2. Background Details

As we realize that local features are the image patterns which vary from its immediate neighbourhood. It is usually influenced by the change of an image property or several properties simultaneously, in spite of the fact that it is not necessary localized precisely on these changes. There are predominantly three properties usually considered- intensity, colour, and texture. Figure 1 shows some examples of local features into account our image as well as in a gray value image. General features can be pointed, in addition edges or small image patches. Ordinarily, some measurements are taken from a region focused on a local feature and changed over into descriptors. The descriptor scan then be used for various applications [1].

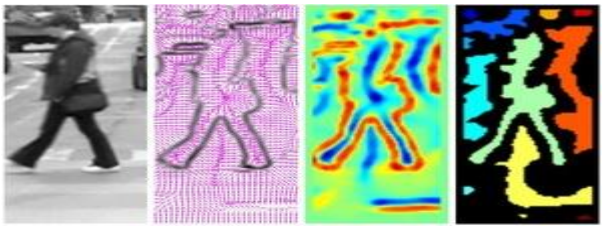

Figure-2: Image Features

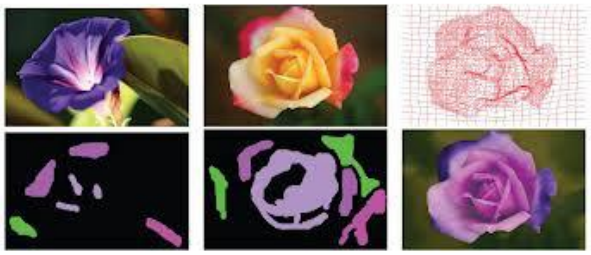

Figure-3: Image Features

Good features should have the following properties:

Distinctiveness/in formativeness: The strength patterns mentioned the discovered features should show a great deal of variety, such that features can be distinguished and matched.

Repeatability: Given two images of the same objector view, taken under diverse viewing conditions, a high rate of the features perceived on part observable in both images should be found in both images.

Locality: The features should be limited, so as to decrease the probability of occlusion and to permit simple model estimates of the geometric and photometric deformations between two images taken under different viewing circumstances (e.g.: based on a local planarity assumption).

Accuracy: The detected features should be accurately localized, in both image location, with respect to scale and possibly shape.

Efficiency: Preferably, the recognition of features in a new image should allow for time-critical applications. Repeatability, debatably the most significant property of all, can be achieved in two different ways: either by invariance or by robustness.

Quantity: The number of detected features should be satisfactorily large, such that a sensible number of features are detected even on small objects. However, the optimal number of features depends on the application. Ideally, the number of identified features should be controllable over a large range by a simple and intuitive threshold. The density of features should reflect the information content of the image to provide a compact image representation.

Robustness: In case of relatively small deformations, it often suffices to make feature detection methods less sensitive to such deformations, i.e., the accuracy of the detection may decrease, but not drastically. Typical deformations that are tackled using robustness are image noise, discretization effects, compression artifacts, blur, etc. Also geometric and photometric deviations from the mathematical model used to obtain invariance are often overcome by including more robustness.

Invariance: When large deformations are to be projected, the preferred approach is to model these mathematical if possible, and then develop methods for feature detection that are unaffected by these mathematical transformations.

2.1 K-Nearest Neighbor (K-NN)

In order to find the user query relevance documents from the keywords further the KNN algorithm is implemented for accurate data listing. The KNN algorithm is stated as [25]:The K-nearest-neighbor algorithm measures the distance between a query scenario and a set of scenarios in the data base. The distance between these two scenarios is estimated using a distance function d(x,y), where x, y are scenarios developed through features, like

$$X=\{x1,x2,x3,\ldots\ldots\ldots..\}$$

$$Y=\{y1,y2,y3,\ldots\ldots\ldots..\}$$

The frequently used distance functions are absolute distance measuring using:

$$d_A(x,y) = \sum_{i=1}^{N} |x_i - y_i|$$

And second is Euclidean distance measuring with:

$$d_A(x,y) = \sum_{i=1}^{N} \sqrt{x_i^2} - y_i^2$$

The overall KNN algorithm is running in the following steps:

- Store the output values of the M nearest neighbors to query scenario Q in vector r = {r1,…...,rm} by repeating the following loop M times:

  a. Go to the next scenario Si in the data set, where I is the current iteration within the domain {1……P}
  b. If Q is not set or q < d (q, Si): q >d (q, Si), t =Oi
  c. Loop until we reach the end of the data set.
  d. Store q into vector c and t into vector r.

- Calculate the arithmetic mean output across r as follows:

$$\bar{r} = \frac{1}{M}\sum_{i=1}^{M} r_i$$

- Return r as the output value for the query scenario q+

3. Literature Review

The given section provides the study about the various recent contributions and research efforts that are improving the image retrieval systems. Learning effective feature representations and similarity measures are vital to the recovery presentation of a content-based image retrieval (CBIR) system. Despite extensive research efforts for decades, it remains one of the most challenging open problems that considerably hinder the successes of real-world CBIR systems. The key challenge has been element to the recognized "semantic gap" matter that survives among low-level image pixels captured by apparatuses and elevated semantic notions apparent by human. Between various techniques, machine learning has been actively investigated as a possible direction to bridge the semantic gap in the long term. Inspired by recent successes of deep learning techniques for computer vision and other applications,

In this paper, **Ji Wan et al [8]** attempt to address an open problem: if deep learning is a hope for bridging the semantic gap in CBIR and how much improvements in CBIR tasks can be achieved by exploring the state-of-the-art deep learning techniques for learning feature representations and similarity measures. Specifically, they investigate a framework of deep learning with application to CBIR tasks with an extensive set of empirical studies by examining a state-of-the-art deep learning method From empirical studies, author find some encouraging results and summarizes some important insights for future research.

Kai Chen et al [9] present the valuation of a creation detection task using the LIRe being and waves (Speeded-Up stout Features) for content-based image retrieval (CBIR). The costing is perform on the Fribourg Product Image Database (FPID) that contains more than 3'000 pictures of consumer products taken use mobile phone cameras in sensible situation. Using the evaluation protocol proposed with FPID, authors explore the performance of different pre-processing and feature extraction. They observe that by using SURF, method can improve significantly the recital on this task. Image resizing and Lucene indexing are used in position to speed up CBIR task with SURF. Also show the benefit of using simple preprocessing of the images such as a comparative crop of the images. The experiments demonstrate the effectiveness of the given method for the product identification task. Content Based Image Retrieval (CBIR) is an important step in addressing image storage and management problems. Latest image skill improvements along with the Internet growth have led to a huge sum of digital disc through the current decades. Different methods, algorithms and system have been future to solve these problems. Such studies revealed the indexing and rescue concept, which have more evolve to Content-Based Image Retrieval. CBIR systems often consider image satisfied via the alleged low-level features for indexing and retrieval, such as color, quality and shape. In order to complete much higher semantic performance, latest systems seek to join low-level with high-level features that contain perceptual information for human. Purpose of this review is to identify the set of methods that have been used for CBR and also to argue various of the key donations in the present decade related to image retrieval and main challenges involved in the adaptation of existing image retrieval technique to construct useful systems that can grip real-world data. By making use of various CBIR approaches accurate, repeatable, quantitative data must be efficiently extracted in order to improve the retrieval accuracy of content-based image retrieval systems.

SatishTunga et al [10], various approaches of CBIR and available algorithms are reviewed. Comparative results (Convolutional Neural Networks) for CBIR tasks under various settings. of various techniques are presented and their advantages, disadvantages and limitations are discussed. The CBIR tend to file and rescue images based on their chart content. CBIR stay left from many evils connected with conformist ways of regain images by keywords. Thus, an increasing concern in the locale of CBIR has been known in recent years. The agreement of a CBIR system largely depends on the exacting image illustration and comparison same function work. The CBIR tends to file and rescue images based on their chart content. CBIR avoid many troubles related with usual ways of retrieve images by keywords. Thus, a growing interest in the area of CBIR has been established in current years. The concert of a CBIR system mainly depends on the exacting image depiction and comparison same function in work.

So a new CBIR system is planned by **Sandeep Singh et al [11]** which will offer correct results as contrast to prior urban system. Soft system will be used in this system. Based Image healing system which evaluate the connection of each image in its data build up to a query image in terms of various visual skin texture and go back the image with beloved range of connection. To expand and put into practice an proficient element origin NN and SVM to extort features according to data set use Auto compute the feature weight by neural network. The accuracy and recall chart in gui according to the get back inside of the images from the datasets. To concern back breeding or feed forward algorithm for neural network classification and calculate cross relationship and apply weakening model for feature matching. The advent of cloud datacenters augments the competence of online data storage. Since huge data is deposited in datacenters, it is essential to efficiently locate and admittance attention data in such a dispersed system. However, customary search methods only permit users to exploration images completed exact-match keywords complete a centralized directory. These methods cannot fulfill the necessities of CBIR.

In this paper, **Jianxin Liao et al [12]** suggest a scalable image recovery structure which can competently support content resemblance exploration and semantic search in the dispersed milieu. Its key notion is to assimilate image feature vectors into distributed hash tables (DHTs) by manipulating the property of locality sensitive hashing (LSH). Thus, images with like content are most probable collected into the similar knot without the information of any worldwide information. For searching semantically close images, the significance response is adopted in our arrangement to overcome the gap between low-level types and elevated types. Author's show that given approach produces high recall rate with good load balancing and only require a few number of hops.

Hanwang Zhang et al [13] present a novel attribute-augmented semantic hierarchy (A2SH) and demonstrate its effectiveness in bridging both the semantic and intention gaps in content based image retrieval (CBIR). A2SHorganizes semantic concepts into multiple semantic levels and augments each concept with a set of related attributes. The attributes are used to describe the multiple facts of the concept and act as the intermediate bridge connecting the concept and low-level visual content. A hierarchical semantic similarity function is learned to characterize the semantic similarities among images for retrieval. To better capture user search intent, a hybrid feedback mechanism is developed, which collects hybrid feedback on attributes and images. This feedback is then used to refine the search results based on A2SH. They use A2SH as a basis to develop a unified content-based image retrieval system. We conduct extensive experiments on a large-scale data set of over one million Web images. Experimental results show that the proposed A2SH can characterize the semantic affinities among images accurately and can shape user search intent quickly, leading to more accurate search results as compared to state-of-the-art CBIR solutions. Effective and efficient image recovery methods have developed the fast and vigorous investigation area since of volatile usage of numerical images. User communication in CBIR scheme involves of a query creation. The user has difficulties in assertion of a query by different schemes which have been introduced in literature.

In this paper, **SunkariMadhu [14]**analyze the two retrieval methods, query by texture and query by color .Texture features involves the invariant histogram characteristics to retrieve the images and color features carry the color histogram in RGB color space to retrieve the images. It is observed from the experimental results, that the query by texture is more effective than the query by color for retrieving the general images. Automatic scrutiny of his to pathological images has been usually utilized leveraging computational image-processing methods and current device knowledge technique. Both CAD (computer-aided diagnosis) and CBIR(content-based image-retrieval) schemes have been fruitfully developed for diagnosis, disease finding, and conclusion maintain in this area. of late, with the rising amount of annotate remedial data, large-scale and data-driven method have emerge to offer a secure of bridging the semantic gap involving images and analytic information.

In this paper, **Xiaofan Zhang et al [15]** center on mounting scalable image retrieval techniques to cope smartly with gigantic histopathological images. specially, they there a direct core hash technique which leverages a small amount of supervise information in learn to compact a 10 000-dimensional pictures feature course into solitary tens of dual bit with the useful signature preserved. These dual codes are next indexed into a hash desk that enables coincident recovery of pictures in a huge database. Critically, the supervise in order is employ to conduit the semantic gap amid low-level image features and top diagnostic information. Author build a scalable image-retrieval frame work based on the supervise hash skill and confirm its performance on some thousand histopathological images acquired from breast atomic tissues. General evaluation are carried out in conditions of image sorting (i.e., benign versus actionable categorization) and rescue test. Agenda achieve about88.1% arrangement truth as well as promise time efficiency. For example, the structure can perform about 800 queries in only 0.01 s, compare favorably with other normally used dimensionality fall and element collection method. Content-based image rescue is a very main area of investigate today. Content Based image rescue (CBIR) is a skill which uses chart features of image such as color, shape, quality, etc. CBIR technology presents a process to get images in large databases by use sole descriptors from a skilled image. A lots of research works had been completed in the past decade to design efficient image retrieval techniques from the image or multimedia databases. Big number of rescue technique has been introduced, but there is no generally usual element mining and rescue skill to be held.

In this paper, **Ashwani Kr. Yadav et al [16]**there a reading of a variety of content-based image rescue system and their behavior, quality scrutiny and a variety of element removal with depiction. In web-scale double recovery, the most real scheme is to total local descriptors into a high dimensionality sign and then decrease it to a small dimensionality. Thanks to this scheme, web-scale image databases can be signified with small directory and discovered using fast pictorial resemblances. However, the calculation of this directory has a very great difficulty, because of the great dimensionality of signature projectors. In this work,

RomainNegrel et al [17] suggest a novel well-organized technique to importantly decrease the sign dimensionality with low computational and storage costs. Our technique is based on the linear projection of the signature onto a small subspace using a sparse forecast matrix. They report numerous trial consequences on two normal datasets (Inria Holidays and Oxford) and with 100kimage distracters. They demonstrate that known technique decreases both the projectors storing cost and the computational cost of projection stage while experiencing a very minor harm in mAP (mean Average Precision) presentation of these calculated signatures.

4. Proposed System

There were several techniques for image retrieval as explained in background section of the paper. Text based Image retrieval system (TIRS) were commonly used systems for feature extraction followed by Content Based Image retrieval system (CIRS) were more advantageous and higher accuracy rate of the extraction of the features from the images. CIRS was the revolutionary technique in feature extraction brings into the market that additionally compresses the image and calculation of pixel values from the images. Greater efficiency of the algorithm is added advantage in the content based system as shown in fig. 6 below:

In this paper, we are proposing a feature extraction and image retrieval system using k-Nearest Neighbors algorithm(k-NN).In this work we investigated a method that prepares a database of features extracted from images provided to the system as an input from the local repository. Images were attached with a tag that also being stored in a database so as to compare with other images. The above process is the part of training the system. Firing test query on the features of the images retrieves the other images with the same tags stored in the database and Levenshtien distance being also calculated and given as an input to k-NN. k-Nearest Neighbors algorithm(k-NN) algorithms is power full algorithm that can be very help full in image extraction. In this process it extracts color features and age features of the images. Results of k-NN algorithm are very useful and accurate in the various applications. An efficient & detailed digital image feature extraction method using k-NN algorithm and image tags architecture of the system is shown in the Fig.6 and divided into four different stages or modules; training, testing ,feature extraction, results generation. The process involves different steps of execution as explained below:

Training:

In this process the system is being trained against the input set of digital images. System extracts and stores the features of the given set of images into the images database and features of the images into the feature database as shown in the architecture. Input images are also added with a tag that will in relevance with image. The tags also stored into the database in which it can use to become a large dataset and retrieve more images with similar tags. In this process, we create a base for the k-NN algorithm.

Testing:

In this part one can fire a query with any digital image or a text query in the form of tags. The system will fetch any digital image and extract features and relevant tags. The next process is to search similar tags into the stored tags in the database previously and search for the image related to tag extracted. In the other part of this operation one can fire text query onto the image for the calculation of Levenshtien distances in the pixels of the images that also become the input of k-NN algorithm.

Feature Extraction:

In this process k-NN comes into the system, it takes features and tags of the queried image also with Levenshtien distances and produce features of the images. We can get a huge amount of features from the system that can be used in various applications.

Results:

The algorithm is so efficient that produces an enormous amount of features including Colour and Age features. The algorithms produce 9*9 colour features that are 81 features for the image. For the Age features it creates 360 degree orientation with 10 numbers of parts that evaluated with 10 parts of 36 features each. The results can be efficiently used in the various applications like face recognition, biometric authentication etc.

The proposed system is found to be efficient and accurate as compared to the previous methods found in the literature. The system can give guarantee about the authenticity of the resulted features and found with the great usefulness in industry and other futuristic research work.

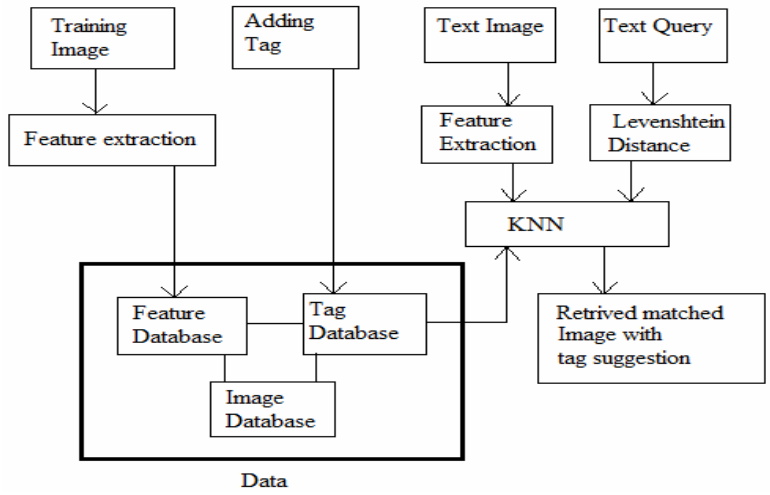

Figure-4: Proposed System of an Efficient & Detailed Digital Image Feature Extraction Method Using K-NN Algorithm.

5. Result Analysis

The implemented enhanced image retrieval technique is evaluated on the basis of the different experimental scenarios and different sets of data. The evaluated performance of the obtained system is described in this chapter with their evaluation and outcomes.

5.1 Precision:

In any data retrieval or search applications the precision is a fraction of search results which is most relevant to the input query. The provided comparative precision of the proposed content based image retrieval system and the available technique [26] are given using figure 5.1. This can be evaluated using the user feedback basis and can be evaluated by the following formula.

$$precision = (releventdocument \cap retrieveddocumets)/retrieveddocuments$$

5.2 Recall:

In data retrieval application or the search application recall values are measured for accuracy measurement in terms of relevant document retrieved or relevant data obtained according to the input user query. This can be evaluated using the following formula.

$$recall = (relevantdoucment \cap retrieveddocuments)/relevantdocuments$$

5.3 F-Measures:

The f-measures of the system demonstrate the fluctuation in the computed performance in terms of precision and recall rates. The f-measures of the system can be approximated using the following formula.

$$F - measures = 2*(precision * recall)/(precision + recall)$$

5.4 Time Consumption:

The amount of time required to complete the retrieval task after providing input to the system is termed as time consumption of the algorithm.

5.5 Memory Used:

The memory used sometimes also called the memory consumption or the space complexity. That amount of main memory required to execute a given algorithm with the amount of data is known as the memory consumption or space complexity of algorithm.

6. Conclusion and Future Work

The implementation of the proposed image retrieval system using the saliency feature extraction for query by image technique is performed successfully. Additionally their performance is also evaluated based on their experimentations some facts are concluded that are reported in this chapter. Furthermore the future extension of the work is also included.

6.1 Conclusion

In this era of technology the need of data and computation is increases continuously. Each and every hand is mounted with the new generation gadgets and smart devices. Additionally these devices are internet enabled thus user continuously search data from the internet and other sources of data. Due to this the need of multimedia contents i.e. image and video contents are also increases by these users. There are two kinds of methods are popular for making search text based techniques and content based techniques. The text based techniques are working on the basis of text and description associated with these multimedia data but the content based techniques are finding the contents that are actually hidden in the multimedia data. Therefore the content based techniques are much effective then the text based techniques.

In this presented work the content based technique is studied in detail and using the available image features i.e. shape, texture and color distribution the images are searched. In addition of that for improving the quality of image search the method is extended with the saliency features computation. That technique helps to group the similar image contents in a group, in addition of the proposed technique is works on the basis query by image technique thus that make more promising outcomes from the retrieval.

References

[1] Jun Huang, Xiaokang Yang, Xiangzhong Fang, Weiyao Lin, and Rui Zhang, "Integrating Visual Saliency and Consistency for Re-Ranking Image Search Results", IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 13, NO. 4, AUGUST 2011

[2] Minsu Cho, Jungmin Lee, and Kyoung Mu Lee,"Reweighted Random Walks for Graph Matching", Computer Vision–ECCV 2010, 2010 – Springer

[3] Priti Maheswary, Dr. Namita Srivastava, "Retrieval of Remote Sensing Images Using Colour & Texture Attribute",(IJCSIS) International Journal of Computer Science and Information Security,Vol. 4,No. 1&2, 2009

[4] S. Oraintara, T. T. Nguyen, "Using Phase and Magnitude Information of the Complex directional Filter Bank for Texture Image Retrieval", IEEE International Conference on Image Processing, Vol. 4, Pages 61-64, Oct. 2007.

[5] P. S. Hiremath and JagadeeshPujari. "Content Based Image Retrieval based on Color, Texture and Shape features using Image and its complement", International Journal of Computer Science and Security, Volume (1) : Issue (4).

[6] Philipp Sandhaus, Mohammad Rabbath, and Susanne Boll,"Employing Aesthetic Principles for Automatic PhotoBookLayout",Advances in Multimedia Modeling, 2011 –Springer.

[7] Michele Saad, "Low-Level Color and Texture Feature Extraction for Content-Based Image Retrieval", Final Project Report, EE 381K: Multi-Dimensional Digital Signal Processing, Pages: 20-28, May 2008

[8] Ji Wan, Dayong Wang, Steven C.H. Hoi, Pengcheng Wu, Jianke Zhu, Yongdong Zhang, Jintao Li, "Deep Learning for Content-Based Image Retrieval: A Comprehensive Study", Copyright 2014 ACM 978-1-4503-3063-3/14/11

[9] Kai Chen and Jean Hennebert, "Content-Based Image Retrieval with LIRe and SURF on a Smartphone-Based Product Image Database", MCPR 2014, LNCS 8495, pp. 231–240, 2014 c Springer International Publishing Switzerland 2014

[10] Satish Tunga, D. Jayadevappa & C. Gururaj, "A Comparative Study of Content Based Image Retrieval Trends and Approaches", International Journal of Image Processing (IJIP), Volume (9) : Issue (3) : 2015

[11] Sandeep Singh, Er.Rachna Rajput, "Content Based Image Retrieval using SVM, NN and KNN Classification", International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 6, June 2015

[12] Jianxin Liao, Di Yang, Tonghong Li, Jingyu Wang, Qi Qi, Xiaomin Zhu, "A scalable approach for content based image retrieval in cloud data center", State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, P.O. Box 296, Beijing 100876, China

[13] Hanwang Zhang, Zheng-Jun Zha, Yang Yang, Shuizheng Yan, YueGao, and Tat- Seng Chua, "Attribute-Augmented Semantic Hierarchy: Towards a Unified Framework for Content-Based Image Retrieval", ACM Trans. Multimedia Comput. Commun.Appl. 11, 1s, Article 21 (September 2014), 21 pages.

[14] Sunkari Madhu, "Content based Image Retrieval: A Quantitative Comparison between Query by Color and Query by Texture", Journal of Industrial and Intelligent Information Vol. 2, No. 2, June 2014
[15] Xiaofan Zhang, Wei Liu, Murat Dundar, Sunil Badve, and Shaoting Zhang, "Towards Large-Scale Histopathological Image Analysis: Hashing-Based Image Retrieval", IEEE Transactions On Medical Imaging, Vol. 34, No. 2, February 2015
[16] Ashwani Kr. Yadav, R. Roy, Vaishali and Archek Praveen Kumar, "Survey on Content-based Image Retrieval and Texture Analysis with Applications", International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 7, No. 6 (2014), pp. 41-50
[17] Romain Negrel, David Picard, Philippe-Henri Gosselin, "Dimensionality Reduction of Visual Features Using Sparse Projectors For Content-Based Image Retrieval", IEEE International Conference on Image Processing, Oct 2014, Paris, France. pp.2192-2196.
[18] Yu-Gang Jiang, Jun Yang, Chong-Wah Ngo, Alexander G. Hauptmann, "Representations of Key point-Based Semantic Concept Detection: A Comprehensive Study", IEEE, 2008
[19] Canny Edge Detection, March 23, 2009
[20] Masoud Nosrati, Ronak Karimi, Mehdi Hariri, "Detecting Circular Shapes From Areal Images Using Median Filter and CHT", World Applied Programming, Vol (2), Issue (1), . 49-54,January 2012
[21] Zhenhua Guo,LeiZhang,David Zhang, "A Completed Modeling of Local Binary Pattern Operator for Texture Classification", IEEE transaction on image processing, 2010
[22] S. Mallat, "A Wavelet Tour of Signal Processing : The Sparse Way", 3 edition, 2009
[23] Joni-Kristian Kamarainen, "Gabor Features in Image Analysis", Machine Vision and Pattern Recognition Laboratory, Lappeenranta University of Technology (LUT Kouvola)
[24] Larnaca, Cyprus, "An MPEG-7 Image Retrieval Systemof Atherosclerotic Carotid Plaque Images", IEEE 12th International Conference on Bioinformatics& Bioengineering (BIBE), 11-13 November 2012
[25] Paul Lammertsma, "K-nearest-neighbor algorithm", paul.luminos.nl/download/document/knn.pdf
[26] M. Yasmin, M. Sharif, I. Irum and S. Mohsin, "An Efficient Content Based Image Retrieval using EI Classification andColor Features", Journal of Applied Research and Technology, Vol. 12, October 2014.

Estimating quality dimensions under technological integration in pharmaceutical supply chain: a fuzzy MCDM approach

Dr. Vinayak Vishwakarma[1]*
Department of Mechanical Engineering.
Acropolis Institute of Technology & Research India
vinayakvishwakarma@gmail.com

Abstract

Being a highly significant part of health care system, developing nation's pharmaceutical industry, supplying medicines across world needed to develop risk averse supply chain to protect global health of human beings and businesses. We have identified eight critical dimensions or categories of quality that can serve as a framework for strategic analysis: Drug performance, Features of drug/ medicine, Reliability of drug/ medicine, Conformance to action requirement, Self-life/usability, Post sales services, Packaging, and Perceived quality. This paper proposes multi criteria decision making (MCDM) methodology based on fuzzy AHP approach to prioritize and rank quality dimensions which are affected by technological integration in PSC. This approach is also suitable in analysing expert's judgments and uncertainty involved in the process of risk assessment. This study identifies eight quality dimensions for investigation as measures of Indian Pharmaceutical supply chain performance through relevant literature and experts opinions. The results of study would be beneficial for industries in managing and reducing the consequences of the poor quality focus in PSC.

1. Introduction

Product quality is rapidly becoming an important competitive issue for the world and late slips by in India are bringing about concern, India is the second-biggest exporter of over-the-counter and physician recommended medications to the United States. Also, the nation is going under expanded investigation by American controllers for wellbeing breaches, misrepresented medication test outcomes and offering fake medications. To help enhance consumer loyalty, more prominent attention is given to the part of value in the store network. This paper will help you see a portion of the quality issues and strategies that are pertinent in current supply chain management. Quality of medicines is growing concern of the world and supply chain management plays a critical role especially to enhance the quality of medicines. This paper is an effort is being made to understand and investigate issues under dimension of quality manufacturing.

Due to rapid growth of human population across corners of the world a rising concerns of counterfeiting and poor quality of medicines are critical for drug industry. Globalization of industry with discrepancy in nation's health care systems with inadequate detection system and interventions causing a supply chain incapable to improve the quality of medicines. Medicines can be classified with respect to quality of compositions often termed as substandard and counterfeited drugs/medicines. The supply chain comprises of various stages of manufacturing lacking in terms poor formulations, old technology production machinery usage ,poor manufacturing, untrained work force , insufficient research and infrastructure leading to substandard medicines. On the other hand counterfeited drugs are intentionally produces to gain undesired financial gains which harms health of business and society.

2. Literature review

There are some new practices that are recently applied to the pharmaceutical industry though they are widely applied in non-pharmaceutical industries, such as: the lean manufacturing; the Six Sigma; the total quality management. They are to enhance the quality of medicinal production. As we are focusing on our objective of integration of such technology to effect on any quality of medicine thorough the dimensions of quality in supply chain.

Table-1: Literature support

S No.	Pharmaceutical supply chain Quality Dimensions	Supported Literature by Authors
1	Drug performance	[1], [2]
2	Features of drug/ medicine	[3], [4]
3	Reliability of drug/ medicine	[3], [5]
4	Conformance to action requirement	[6][7]
5	Self-life/usability	[8], [9]
6	Post sales services	[10], [11]
7	Packaging	[6], [12]
8	Perceived quality	[13]

The questionnaire has been designed to investigate the quality dimensions effecting quality production of medicine. The amount supply chain integration has directly affects the various dimension of quality production. Here respondents asked to rate the integration level with respect to all quality dimensions.

3. Research methodology

The motive behind selection of methodology is the identification, evaluation and prioritization of identified quality dimensions. The decision makers include experts (Domain specialist, Supply chain managers, and senior executive) from two major industrial region of country. In this research 8 QDs of PSCM adoption comprising qualitative and quantitative, recognized through literature and discussion (See Table1). This study utilizes Fuzzy Analytical Hierarchical Process to rank and evaluate the identified QDs for successful integration with PSCM practices.

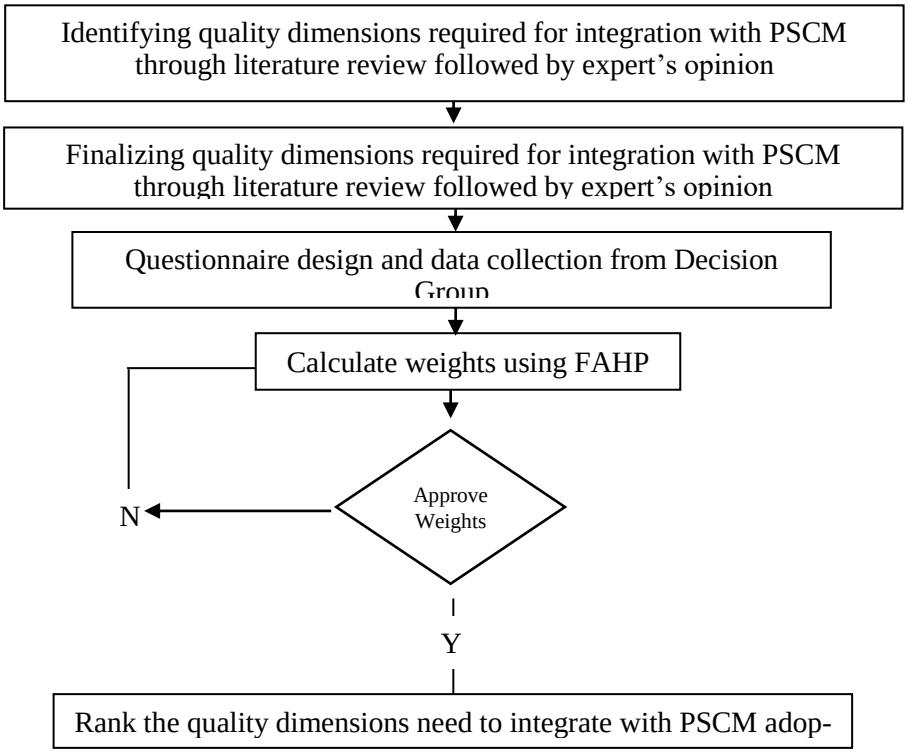

Figure-1: Proposed FAHP methodology for quality dimensions assessment

4.1 Fuzzy AHP

AHP approach introduced by Satty (1980); it was a numerical approach of MCDM. The application of AHP has some limitations in certain environment; so that create need to have Fuzzy methodology to answer such problem. The fuzzy AHP approach includes uncertainty and vagueness of expert's judgments in linguistic variables. Recently this approach is used by (Prakash et al.; 2014, 2015c, d) and (Prakash & Barua 2015a; b; e, Vishwakarma et al; 2015)

Table-2: Assigned TFN

Linguistic variables	Assigned TFN
Equal	(1, 1, 1)
Very Low	(1, 2, 3)
Low	(2, 3, 4)
Medium	(3, 4, 5)
High	(4, 5, 6)
Very High	(5, 6, 7)
Excellent	(7, 8, 9)

Chang's extent analysis (1996) is the description of FAHP process; the values of extent method for each criterion are determined using following notation.

$M_{g_i}^1, M_{g_i}^2, M_{g_i}^3 \ldots\ldots, M_{g_i}^m$ Where g_i is the goal set (i = 1, 2, 3, 4, 5 …....n) and all the $M_{g_i}^j$(j = 1, 2, 3, 4, 5 ....., m) are TFNs. The steps of Chang's analysis can be given as in the following:

Step 1: The fuzzy synthetic extent value (S_i) with respect to the i^{th} criterion is defined as,

$$S_i = \sum_{j=1}^{m} M_{g_i}^j \times \left[\sum_{i=1}^{n}\sum_{j=1}^{m} M_{g_i}^j\right]^{-1} \qquad \text{.....……..(3.10)}$$

$$\sum_{j=1}^{m} M_{g_i}^j = \left(\sum_{j=1}^{m} l_{ij}, \sum_{j=1}^{m} m_{ij}, \sum_{j=1}^{m} u_{ij}\right)$$

$$\left[\sum_{i=1}^{n}\sum_{j=1}^{m} M_{g_i}^j\right]^{-1} = \left(\frac{1}{\sum_{i=1}^{n}\sum_{j=1}^{m} u_{ij}}, \frac{1}{\sum_{i=1}^{n}\sum_{j=1}^{m} m_{ij}}, \frac{1}{\sum_{i=1}^{n}\sum_{j=1}^{m} l_{ij}}\right)$$

Where l is the lower limit value, m is the most promising value and u is the upper limit value.

Step 2: The degree of possibility of

$S_2 = (l_2, m_2, u_2) \geq S_1 = (l_1, m_1, u_1)$ is defined as below

$V(S_2 \geq S_1) = \sup_{y \geq x}[\min(\mu_{S_1}(x), \mu_{S_2}(y)]$

and x and y are the values on the axis of membership function of each criterion.

This expression can be equivalently written as given in equation 3.11 below:

$$V(S_2 \geq S_1) = \begin{cases} 1 & \text{if } m_2 \geq m_1 \\ 0 & \text{if } l_1 \geq u_2 \end{cases}$$

$$\frac{l_1-u_2}{(m_2-u_2)-(m_1-l_1)}=\mu d, \quad \text{otherwise} \qquad \ldots\ldots\ldots (3.11)$$

Where μd is the highest intersection pointμ_{S_1} and μ_{S_2} (see figure 3.5)

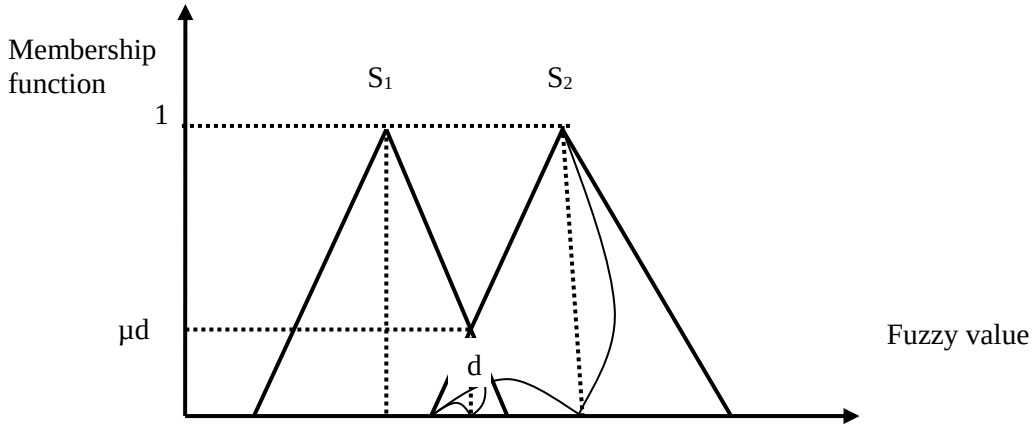

Figure-2: The intersection of fuzzy numbers

To compare S_1 and S_2 we need both $V(S_1 \geq S_2)$ and $V(S_2 \geq S_1)$.

Step 3: The degree of possibility for a convex fuzzy number S to be greater than *k* convex fuzzy numbers S_i (*i*= 1,2,.....,*k*) can be defined by

$V(S \geq S_1, S_2, \ldots\ldots, S_k) = \min V(S \geq S_i),\ i = 1,2,\ldots\ldots,k$

Assume that $d'(A_i) = \min V(S_i \geq S_k)$ ……...(3.12)

For $k = 1, 2, \ldots, n$, $k \neq i$, Than the weight vectors are given in equation 3.13 as,

$W' = (d'(A_1), d'(A_2), \ldots\ldots, d'(A_m))^T$ ……..(3.13)

Step 4: Via normalization, the normalized weight vectors are given in equation 3.14 as,

$W = (d(A_1), d(A_2), \ldots\ldots, d(A_m))^T$ ……(3.14)

4.2 *Calculation of the Value of Fuzzy Synthetic Extent*

Decision group has to make pair-wise comparison of QDs, defined by TFN as given in table 2. The TFN comparison matrix of the QDs is given in table 3.

Calculations of the fuzzy synthetic extent of 8 QDs are given below by using eq. (3.10).

S (QD1) = $(8.4, 13.5, 18.67) \otimes [74.08, 104.67, 138.84]^{-1}$
= $(0.06, 0.129, 0.251)$
S (QD2) = $(10.78, 15.08, 19.83) \otimes [74.08, 104.67, 138.84]^{-1}$
= $(0.076, 0.144, 0.267)$
S (QD3) = $(9.92, 14.33, 19.5) \otimes [74.08, 104.67, 138.84]^{-1}$
= $(0.071, 0.136, 0.263)$
S (QD4) = $(12.78, 17.08, 21.83) \otimes [74.08, 104.67, 138.84]^{-1}$
= $(0.092, 0.163, 0.294)$
S (QD5) = $(8.23, 10.67, 13.67) \otimes [74.08, 104.67, 138.84]^{-1}$
= $(0.059, 0.101, 0.185)$

$S(QD6) = (8.98, 12.33, 16.17) \otimes [74.08, 104.67, 138.84]^{-1}$
$= (0.064, 0.117, 0.218)$
$S(QD7) = (6.2, 8.58, 11.33) \otimes [74.08, 104.67, 138.84]^{-1}$
$= (0.044, 0.082, 0.152)$
$S(QD8) = (8.78, 13.08, 17.83) \otimes [74.08, 104.67, 138.84]^{-1}$
$= (0.063, 0.125, 0.24)$

V values and minimum degree of possibility are calculated by using the equation 3.11, 3.12 respectively. $m(QD1) = \min V(S_1 \geq S_k) = 0.823$ and other values are m(QD2) = 0.901, m(QD3) = 0.866, m(QD4) = 1, m(QD5) = 0.455, m(QD6) = 0.735, m(QD7) = 0.428, m(QD8) = 0.795

Weight vector is given by:
$W_V = (0.823, 0.901, 0.866, 1, 0.455, 0.735, 0.428, 0.795)^T$
Final weights determined after normalization process-
W = (0.137, 0.150, 0.144, 0.166, 0.075, 0.122, 0.071, 0.132)

These calculations have done by using MS-Excel. The final weights of QDs and the final ranking are presented in table 4.

Table-3: Triangular fuzzy number based pair-wise judgment matrix for quality dimensions

	QD1	QD2	QD3	QD4	QD5	QD6	QD7	QD8
QD1	(1, 1, 1)	(1, 2, 3)	(1, 2, 3)	(0.2, 0.25,0.33)	(2, 3, 4)	(1, 2, 3)	(0.2, 0.25,0.33)	(2, 3, 4)
QD2	(0.33, 0.5, 1)	(1, 1, 1)	(0.25, 0.33,0.5)	(0.2, 0.25,0.33)	(3, 4, 5)	(3, 4, 5)	(2, 3, 4)	(1, 2, 3)
QD3	(0.33, 0.5, 1)	(2, 3, 4)	(1, 1, 1)	(3, 4, 5)	(1, 2, 3)	(2, 3, 4)	(0.33, 0.5, 1)	(0.25, 0.33,0.5)
QD4	(3, 4, 5)	(3, 4, 5)	(0.2, 0.25,0.33)	(1, 1, 1)	(3, 4, 5)	(0.25, 0.33,0.5)	(2, 3, 4)	(0.33, 0.5, 1)
QD5	(0.25, 0.33,0.5)	(0.2, 0.25,0.33)	(0.33, 0.5, 1)	(0.2, 0.25,0.33)	(1, 1, 1)	(3, 4, 5)	(3, 4, 5)	(0.25, 0.33,0.5)
QD6	(0.33, 0.5, 1)	(0.2, 0.25,0.33)	(0.25, 0.33,0.5)	(2, 3, 4)	(0.2, 0.25,0.33)	(1, 1, 1)	(2, 3, 4)	(3, 4, 5)
QD7	(3, 4, 5)	(0.25, 0.33,0.5)	(1, 2, 3)	(0.25, 0.33,0.5)	(0.2, 0.25,0.33)	(0.25, 0.33,0.5)	(1, 1, 1)	(0.25, 0.33,0.5)
QD8	(0.25, 0.33,0.5)	(0.33, 0.5, 1)	(2, 3, 4)	(1, 2, 3)	(2, 3, 4)	(0.2, 0.25,0.33)	(2, 3, 4)	(1, 1, 1)

Source: Fuzzy AHP Analysis

Table-4: Final ranking of quality dimensions with integration PSCM

S. No.	Critical factors	Code	Final weights	Rank
1	Drug performance	QD1	0.13709	4
2	Features of drug/ medicine	QD2	0.15012	2
3	Reliability of drug/ medicine	QD3	0.14430	3
4	Conformance to action requirement	QD4	0.16645	1
5	Self-life/usability	QD5	0.07590	7
6	Post sales services	QD6	0.12241	6
7	Packaging	QD7	0.07133	8
8	Perceived quality	QD8	0.13241	5

Source: Fuzzy AHP Analysis

5. Results and Discussions

In this outlook, the present study tries to add in the literature of PSC; through identifying, finalizing, and prioritizing the quality dimensions related to PSC, so they can manage on strategic level in a business. The findings of this study would be useful for industry and management to become more capable in analyzing the PSC quality dimensions and reduce their consequences. The identified quality dimensions were analyses to priority using fuzzy AHP approach. The analysis of data by fuzzy AHP will assist managers to overcome the problem of human subjectivity in analyzing

the PSC quality dimensions. It will help to provide a measure determining the relative concerns of recognized categories of quality dimensions and specific quality dimensions in PSC. The priority wise concern for the identified eight categories of quality dimensions is given as, QD4>QD2>QD3>QD1>QD8>QD6>QD5>QD7 i.e., Conformance to action requirement as quality dimensions are the most important and need a greater managerial concern as compared to other categories of quality dimensions in enhancing the PSC performance. The priorities for specific quality dimensions were also derived. At last, sensitivity analysis was performed to test the stability of priority ranking for the finalized category of quality dimensions and specific quality dimensions.

7. Limitations of the study and scope of future work

We have used fuzzy AHP approach for prioritizing PSC quality dimensions to improve the performance of Indian pharmaceutical industry. All pair comparisons in fuzzy AHP have been assigned by experts. From the relevant literature and experts views in detail, quality dimensions have been identified and ranked. As it is natural, views of decision makers may be subjective and vary. Different MCDM approaches may be applied using several approaches such as ANP, VIKOR, MAUT, DEMATEL and TOPSIS for the similar problem and outcomes/results can be matched in the further studies.

References

[1] R. P. Cogdill and J. K. Drennen, "Risk-based quality by design (QbD): A Taguchi perspective on the assessment of product qulity, and the quantitative linkage of drug product parameters and clinical performance," J.Pharm. Innov., vol. 3, no. 1, pp. 23–29, 2008

[2] X. Yu, C. Li, Y. Shi, and M. Yu, "Pharmaceutical supply chain in China: current issues and implications for health system reform", Health policy, vol 97, pp. 8-15, sep 2010

[3] L. Schweizer, "Knowledge transfer and R&D in pharmaceutical companies: A case study," J. Eng. Technol. Ma nag. JETM, vol. 22, no. 4, pp. 315–331, 2005.

[4] P. N. Newton, M. D. Green, and F. M. Fernández, "Impact of poor-quality medicines in the 'developing'world", Trends Phamacol. SCI. vol. 31, no 3, pp. 99-101, 2010.

[5] J. N. Sangshetti, M. Deshpande, R. Arote, Z. Zaheer, and D. B. Shinde, "Quality by design approach: Regulatory need," Arab. J. Chem., 2014.

[6] L Kubitza, "Pharma 2020: Marketing the future -Which path will you take?" PWC, pp. 1–32, 2009.

[7] S. Görög, "Drug safety, drug quality, drug analysis", J. Pharm. Biomed. Anal., vol. 48, no. 2, pp. 247– 253, 2008.

[8] R. Narayan, "Universal health care in India: Missing core determinants," Lancet, vol. 377, no. 9769, pp. 883–885, 2011.

[9] Kubitza, "Pharma 2020: Challenging business models - Which path will you take?" ,PWC, no. 30, pp. 1–24, 2009.

[10] H. Kurata and S. H. Nam, "After-sales service competition in a supply chain: Optimization of customer satisfaction level or profit or both?," Int. J. Prod. Econ., vol. 127, no. 1, pp. 136–146, 2010.

[11] D. DelVecchio, D. H. Henard, and T. H. Freling, "The effect of sales promotion on post-promotion brand Preference: A meta-analysis," J. Retail., vol. 82, no. 3, pp. 203–213, 2006.

[12] S. A. Narayana, R. Kumar Pati, and P. Vrat, "Managerial research on the pharmaceutical supply chain a critical review and some insights for future directions," J. Purch. Supply Manag., vol. 20, no. 1, pp. 18–40, Mar. 2014.

[13] C. Sánchez-Rodríguez and Á. R. Martínez-Lorente, "Quality management practices in the purchasing function: An empirical study", International Journal of operation and production management, vol 24, no. 7. pp. 666-687, 2004.