A Review on the Evaluation of Performance towards Data Mining Algorithms

Monelli Ayyavaraiah[1], J. Vahini Prasanthi[2], Humera Shadan[3]
[1]Assistant Professor, Department of Information Technology, MGIT, Hyderabad
[2]Assistant Professor, Department of CSE, MGIT, Hyderabad
[3]Assistant Professor, Department of CSE, MGIT, Hyderabad

ABSTRACT: Data mining has turned out to be a standout amongst the most fundamental devices in various fields. The increments in data sizes and algorithmic complexities require the computational energy of chip to increment much further. In this paper, we introduce nitty gritty qualities from the hardware and software points of view for an arrangement of delegate data mining programs. We first outline MineBench, a benchmarking suite containing agent data mining applications from various classes including two classification, two association rule mining, and four clustering applications. We assess the MineBench applications on a 8-way Shared Memory Parallel (SMP) machine and dissect their critical performance attributes. Amid the evaluation, the info datasets and the quantity of processors utilized are shifted to gauge the scalability of the applications in our benchmark suite. We introduce the outcomes in light of attributes, for example, scalability, I/O intricacy, portion of time spent in the OS mode, and communication/synchronization overheads. This data can help creators of future frameworks and in addition software engineers of new data mining calculations to accomplish better framework and algorithmic performance.

Index Terms: parallel computing, data mining, Performance evaluation, benchmark

1. Introduction

As the data sizes amassed from different fields are exponentially expanding, data mining strategies that concentrate data from immense measure of data have turned out to be prevalent in business and logical areas, including advertising, client relationship administration, scoring and hazard administration, suggestion frameworks, misrepresentation recognition, cosmology recreation, atmosphere demonstrating, bioinformatics, medicate revelation, interruption location, World Wide Web, and so forth [1].

As the measure of data gathered expands, we should use considerably more entangled data mining applications. In any case, the performance of PC frameworks is enhancing at a slower rate contrasted with the expansion in demand for data mining applications. Late patterns recommend that the framework performance (data in view of memory and I/O bound workloads like TPC-H) has been enhancing at a rate of 10-15% every year, though, the volume of data that is gathered copies each year. Having watched this pattern, analysts have concentrated on effective executions of various data mining calculations. Among these, a noteworthy approach taken is the advancement of parallel and dispersed renditions of such calculations. While these calculations have been proficiently enhanced, the fundamental qualities that characterize these calculations stay under contemplated. Such data thusly can be used amid the execution of the calculations and the plan/setup of the computing frameworks. Understanding the performance bottlenecks is basic not just for processor fashioners to adjust their structures to data mining applications, yet in addition for developers to adjust their calculations to the modified necessities of applications and designs.

In this paper, we attempt to explore data mining applications to recognize their qualities in a successive and a parallel execution condition. We initially set up a benchmarking suite of applications, called MineBench, that include calculations usually utilized as a part of data mining. We trust that dissecting the conduct of an entire benchmarking suite will absolutely give a superior understanding of the hidden bottlenecks for data mining applications. Such examinations have been improved the situation different fields, for example, SPEC [2], TPC [3], SPLASH [4], MediaBench [5] are outstanding benchmarks worked through such an exertion. At that point, we break down the design properties of these applications, and likewise consider their scalabilty regarding these attributes. We break down the qualities of the applications in Shared Memory Parallel (SMP) machine. Notwithstanding their constraint on scalability, SMPs have turned into the most widely recognized parallel computing write in industry because of their effortlessness. By breaking down the application attributes in this delegate multiprocessor framework, we give an understanding into the parallel applications, which can be conceivably useful when creating parallel data mining calculations on SMPs.

Whatever remains of the paper is sorted out as takes after. In the following section, we outline the related work. In Section 3, we talk about the data mining applications that are incorporated into our benchmarking suite. Section 4 displays the evaluation strategy. The qualities of our benchmark applications are introduced in Section 5. Section 6 condenses the outcomes.

2. Related Work

An incredible number of productive data mining advancements have been created in the current years [1, 6]. With a specific end goal to take care of the demand in performance, numerous parallel data mining calculations have likewise been produced. In the previous decade, most research on parallel data mining [1, 7, 8] has been centered around Distributed Memory Parallel machines because of its ability for gigantic parallelism. Nonetheless, Shared Memory Parallel machines are turning into the overwhelming kinds of parallel machines in industry due to its effortlessness and low to medium level of parallelism other than its ostensible cost. A couple of parallel calculations on SMPs have been proposed in [9, 8].

Performance portrayal of individual data mining calculation has been done in [14, 15], where they center around the memory and reserve practices of a choice tree acceptance program. Nonetheless, we trust that examining the practices of a total data mining benchmarking suite will unquestionably give a superior understanding of the fundamental bottlenecks for data mining applications.

3. MineBench Application Suite

We initially set up MineBench, a benchmarking suite containing data mining applications. The determination rule is to incorporate classes and applications that are normally utilized as a part of industry and are probably going to be utilized as a part without bounds, subsequently accomplishing a sensible portrayal of the current applications. MineBench can be utilized by the two software engineers and processor planners for effective framework outline. The applications and in addition imperative attributes of the applications are recorded in Table 1, which displays the applications, the classification they have a place with, a short depiction of the applications, and the programming dialect used to execute it.

3.1 Classification Programs

A classification calculation is to utilize a preparation dataset to fabricate a model with the end goal that the model can be utilized to relegate unclassified records into one of the characterized classes. Classification has applications in differing fields, for example, retail promoting, misrepresentation discovery, and outline of telecommunication benefit designs [1].

Table 1. Applications of MineBench.

Algorithms	Category	Description	Lang.
ScalParC	Classification	Decision tree classifier	C
Naïve Bayesian	Classification	Statistical classifier based on class conditional independence	C++
K-means	Clustering	Partitioning method	C
Fuzzy K- means	Clustering	Fuzzy logic based K-means	C
BIRCH	Clustering	Hierarchical method	C++
HOP	Clustering	Density-based method	C
Apriori	ARM	Horizontal database, level- wise mining based on Apriori property	C/C++
Eclat	ARM	Vertical database, break large search space into equivalence class	C++

ScalParC is a productive and versatile variety of choice tree classification [7]. The choice tree show is worked by recursively part the preparation dataset in light of an ideal standard until the point that all records having a place with every one of the allotments bear a similar class name. Among numerous classification strategies proposed throughout the years, choice trees are especially suited for data mining, since they are manufactured moderately quick contrasted with different techniques, getting comparative or regularly better precision [16], and simple to decipher [17].

Bayesian classifiers are factual classifiers in view of Bayes' Theorem. They foresee the likelihood that a record has a place with a specific class. A basic Bayesian classifier, called Naive Bayesian classifier [18], is similar in performance to choice trees and shows high precision and speed when connected to vast databases.

3.2 Clustering Programs

Clustering is the way toward finding the gatherings of comparative objects from a database to describe the hidden data dispersion. It has wide applications in market or client division, design acknowledgment, organic investigations, and spatial data examination [1]. For the most part, clustering algorithms can be arranged into four classifications: parceling based, various leveled based, thickness based, and matrix based.

K-means [19] is a parcel based strategy and is seemingly the most regularly utilized clustering procedure. Given the client gave parameter k, the underlying k group focuses are randomly chosen from the database. At that point, K-means allocates each protest its closest bunch fixate in view of some closeness work. Once the assignments are finished, new focuses are found by the mean of the considerable number of objects in each bunch. This procedure is rehashed until the point that two back to back emphasess produce a similar bunch task.

K-means doles out a data protest be not to be an individual from a specific bunch. The Fuzzy K-means algorithm [20] unwinds this condition by expecting that a data question can have a likelihood of participation in each bunch. For each protest, the aggregate of the probabilities to all clusters equivalents to 1. Contrasted with K-means, the count for the fuzzy enrollment brings about higher computational cost. Notwithstanding, the adaptability of doling out objects to different clusters may be important to produce better clustering characteristics.

BIRCH [21] is a various leveled clustering technique that utilizes a progressive tree to speak to the closeness of data objects. BIRCH first sweeps the database to construct a clustering-include (CF) tree to condense the bunch portrayal. At that point, a chose clustering algorithm, for example, K-means, is connected to the leaf nodes of the CF tree. For a vast database, BIRCH can accomplish great performance and scalability. It is additionally viable for incremental clustering of approaching data objects.

Thickness based strategies develop clusters as indicated by the thickness of neighboring objects or as indicated by some other thickness work. Bounce [22], initially proposed in astronomy, is an average thickness based clustering strategy. In the wake of doling out an estimation of its thickness for every molecule, HOP connects every molecule with its densest neighbor. The task procedure proceeds until the point that the densest neighbor of a molecule is itself. All particles achieving this state are bunched as a gathering. Bounce has diverse applications in molecular biology, geology, and astronomy.

3.3 Association Rule Mining (AMR) Programs

Association rule mining is to locate the arrangement of all subsets of things or qualities that oftentimes happen in database records. ARM can find intriguing association connections among the expansive number of business exchange records. This can help business basic leadership forms, for example, inventory configuration, cross-marketing, and misfortune pioneer investigation [1].

Apriori [23] is ostensibly the most powerful ARM algorithm. It investigates the level-wise mining of Apriori property: all nonempty subsets of an incessant itemset should likewise be visit. At the kth emphasis (for k > 1), it shapes visit (k+1)- itemset candidates in view of the successive k-itemsets and checks the database to locate the entire arrangement of continuous (k+1)- itemsets, Lk+1. To enhance the proficiency, a hash-based strategy is utilized to lessen the measure of the candidate set.

Eclat [8] utilizes a vertical database organize. It can decide the help of any k-itemset by essentially crossing the id-rundown of the initial two (k-1)- subsets that offer a typical prefix. It breaks the pursuit space into little, autonomous, and sensible chunks. Proficient cross section traversal methods are utilized to distinguish all the genuine maximal continuous itemsets.

Parallel Implementation

As said in the before sections that piece of we will likely test the scalability of data mining applications on SMPs, parallel forms of our benchmark applications are additionally given. We incorporate successive (serial/single processor) usage and evaluation comes about for all applications. Trial comes about for 5 parallel applications out of the 8 benchmark applications have been given: ScalParC (Classification), K-means, Fuzzy K-means, HOP (Clustering), and Apriori (ARM). We picked these applications since these parallel algorithms are regularly found in the writing. ScalParC is parallelized on SMPs utilizing the plan introduced in [9]. Basic data parallelism is misused to parallelize K-means, Fuzzy K-means, and HOP. We execute parallel Apriori in light of the Common Candidate Partitioned Database (CCPD) system proposed in [8].

4. Evaluation methodology

We consider every application both from the algorithmic and the framework point of view. Schedules inside every application are broke down in detail both from the useful and engineering granularity to recognize the key parameters in every algorithm.

4.1 Hardware Platform

We direct our evaluation on an Intel IA-32 multiprocessor stage, which comprises of an Intel Xeon 8-way Shared Memory Parallel (SMP) machine running Linux working framework, a 4GB shared memory and 1024 KB L2 reserve for every processor. Every processor has 16KB non-blocking, coordinated L1 guideline and data reserves. The quantity of processors is differed to think about the scalability.

4.2 Software Tools

In every one of the analyses, we utilize VTune Performance Analyzer [24] for profiling the capacities inside our application, and for estimating their breakdown execution times. VTune counter screen gives a wide collection of measurements. We look at changed qualities of the applications: execution time, part of time spent in the OS space, communication/synchronization intricacy, and I/O multifaceted nature.

In parallel executions of the applications, we utilize OpenMP pragmas [25]. OpenMP is a particular for an arrangement of compiler mandates, library schedules, and condition factors that can be utilized to indicate shared memory parallelism. Because of its straightforwardness, OpenMP is quickly getting to be a standout amongst the most generally utilized programming styles for SMPs. In SMPs, processors convey through shared factors in the single memory space. Synchronization is utilized to facilitate forms. VTune gives the total time spent on various sorts of pragmas, with the goal that we can precisely quantify the time spent on synchronization.

4.3 Dataset Characteristics

Info data is a vital piece of the data mining applications. For ScalParC and Naïve Bayesian, we utilize a synthetic dataset, F26-A32-D250K, produced by the IBM Quest data generator [26]. The documentation F26-A32-D250K indicates a dataset with Function 26, Attribute measure 32, and Data including 250,000 records. For Apriori and Eclat. This documentation signifies the dataset contains 2,000,000 exchanges, the normal exchange measure is 20, and the normal size of the maximal potentially vast itemsets is 6. The quantity of things is 1000 and the quantity of maximal potentially substantial itemsets is 2000. We utilize a genuine picture database for K-means and Fuzzy K-means. This database comprises of 17,695 view pictures. Each photo is spoken to by two highlights: shading and edge. We utilize the dataset spoke to by edge in our test. Since the clustering nature of K-means techniques exceptionally relies upon the information parameter k, we perform both K-means with ten diverse k esteems going from 4 to 13. The planning comes about gave in this paper are the accumulated time for the ten runs.

5. Program Characteristics

In this section, we examine a few attributes of MineBench programs. For every trademark, we dissect how the outcomes shift when we change the quantity of processors utilized as a part of the execution. Our measures of intrigue incorporate the general program execution time, the working framework overheads, I/O times and synchronization times. The advantages and drawbacks of utilizing a shared memory demonstrate for our data mining algorithms are additionally examined.

5.1 Execution Time

The best speedup, 6.06 on 8 processors, is seen in ScalParC. The adjusted data segment on to processors limits the memory get to contention for simultaneous read-compose activities on the shared factors. In the event that data is equally appropriated, every processor can work freely (speedier) by getting to just its individual data block in the memory without expecting access to memory blocks of other processors. Jump takes after ScalParC regarding the accomplished speedups. Apriori has impediments when extended to SMPs. This is because of the critical measure of nuclear access to the shared hash-tree structure and the idea of lopsided exchange data.

Table 2. Execution times for applications

Program	P1(s)	P4	P8
HOP	52.7	1.92	6.06
K-means	12.9	3.9	4.96
Fuzzy K- means	146.8	3.44	5.42
BIRCH	31.7	-	-
ScalParC	110.6	3.88	5.12
Bayesian	25.1	-	-
Apriori	102.7	2.66	3.36
Eclat	81.5	-	-

5.2 Operating System Overheads

In any program, the CPU utilization is part into working framework (OS) and client space. The OS overheads incorporate components like framework calls (for process/string administration, invoking locks, handling hardware interferes), and portion of moderate framework cradles amid program execution. In Figure 1, we show the OS part (as a level of aggregate execution time) of every individual application. At the point when the quantity of processors is 1, the working framework overheads are negligible. The greatest overhead (1.7%) on 1 processor is seen for BIRCH. At the point when the quantity of processors sent is expanded, the OS part increments radically because of the parallelization overheads. Under the OpenMP programming condition, each OpenMP (omp) order includes additional cycles of overhead. Program locks (which are fundamentally framework locks) utilized as a part of parallelization likewise add to the OS overheads. On the whole, the more processors, the more OS overheads. Among the applications, K-means has the most noticeably awful overhead, 40%. It clarifies the poor scalability of K-means. This is because of the omp orders and locks amid the parallelization of K-means.

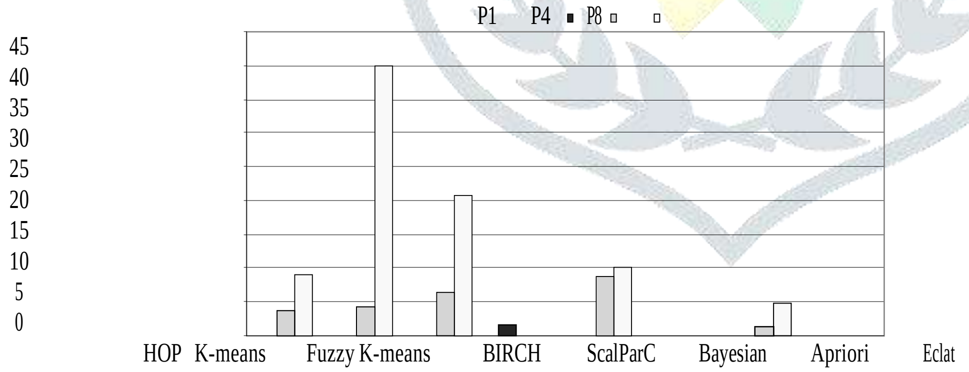

Figure 1. MineBench applications overheads towards OS as a % of total execution time.

6. Conclusion

In this paper, we present and assess MineBench, a benchmarking suite for data mining applications. It contains 8 agent applications: two association rule mining algorithms, two classification algorithms, and four clustering algorithms. We have contemplated vital attributes of the applications when executed on a 8-way SMP machine. Our outcomes show that typically the OS overheads, the synchronization overheads, and the I/O time are generally little in MineBench applications. These outcomes show that upgrades in the performance of processors are likely to significantly affect the general performance of data mining systems. What's more, methods, like prefetching, ought to likewise enhance the performance of the processor impressively. To enhance the performance of their applications, system designers and also programmers can use the qualities of MineBench and accomplish better system performance

References

[1]Han, J. and M. Kamber, Data Mining: Concepts and Techniques (Morgan Kaufmanm, 2001).
[2]Standard, P.E.C., Spec CPU2000: Performance Evaluation in the New Millennium (Version 1.1. December 27, 2000).
[3]TPC, Transaction Processing Performance Council, http://www.tpc.org/.

[4]Woo, S.C., et al., The SPLASH-2 Programs: Characteriazation and methodological contemplations, International Symposium on Computer Architecture, June 1995.
[5] Lee, C., M. Potkonjak, and W.H. Mangione-Smith, MediaBench: A Tool for Evaluating and Synthesizing Multimedia and Communicatons Systems, International Symposium on Microarchitecture, 1997.
[6] Michalski, R.S., I. Brakto, and M. Kubat, Machine Leaning and Data Mining: Methods and Applications (John Wiley and Sons, New York, 1998).
[7] Joshi, M.V., G. Karypis, and V. Kumar, ScalParC: A New Scalable and Efficient Parallel Classification Algorithm for Mining Large Datasets, International Parallel Processing Symposium, 1998.
[8] Zaki, M.J., Parallel and Distributed Association Mining: A Survey, IEEE Concurrency, Special Issue on Parallel Mechanisms for Data Mining, 7(4), Dec. 1999, 14-25.
[9] Zaki, M., C.- T. Ho, and R. Agrawal, Scalable Parallel Classification for Data Mining on Shared-Memory Multiprocessors, IEEE International Conference on Data Engineering, March 1999.
[10] Hankins, R., et al., Scaling and Charaterizing Database Workloads: Bridging the Gap amongst Research and Practice, Intl. Symposium on Microarchitecture, 2003.
[11] Keeton, K., et al., Performance Characterization of a Quad Pentium Pro SMP utilizing OLTP Workloads, Intl. Symposium on Computer Architecture, 1998.
[12] Ranganathan, P., et al., Performance of database workloads on shared-memory system without-of-arrange processors, eighth International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS), San Jose, CA, 1998.
[13] Trancoso, P., et al., The memory performance of DSS business workloads in shared-memory multiprocessors, Third International Symposium on High-Performance Computer Architecture (HPCA), Jan. 1997.
[14] Bradford, J.P. and J. Fortes, Performance and Memory-Access Characterization of Data Mining Applications, Workload Characterization: Methodology and Case Studies, Dallas, TX, Nov. 1998.
[15] Kim, J.- S., X. Qin, and Y. Hsu, Memory Characterization of a Parallel Data Mining Workload, Workshop on Workload Characterization: Methodology and Case Studies, Dallas, TX, Nov. 1998.
[16] Michie, D., D.J. Spiegelhalter, and C.C. Taylor, Machine Learning, Neural and Statistical Classification (Ellis Horwood, 1994).
[17] Quinlan, J., C4.5 Programs for Machine Learning (Morgan Kaufmann, 1993).
[18] Domingos, P. and M. Pazzani, Beyond autonomy: Conditions for optimality of the straightforward Bayesian classifier, thirteenth Internaltional Conference on Machine Learning, 1996.
[19] MacQueen, J, Some Methods for Classification and Analysis of Multivariate Observations, fifth Berkeley Symposium on Mathematical Statistics and Probability, 1967.
[20] Bezdek, J.C., Pattern Recognition with Fuzzy Objective Function Algorithms (Plenum Press, New York, 1981).
[21] Zhang, T., R. Ramakrishnan, and M. Livny, BIRCH: A proficient data clustering technique for extensive databases, SIGMOD, June 1996.
[22] Eisenstein, D.J. and P. Hovel, HOP: A New Group Finding Alogrithm for N-Body Simulations, Journal of Astrophysics, 1998, 498: pp. 137-142.
[23] Agrawal, R., et al., Fast Discovery of Association Rules, Advances in Knowledge Discovery and Data Mining, 1995.
[24] Intel, C., VTune Performance Analyzer, http://www.intel.com/software/items/vtune/.
[25] OpenMP, OpenMP: Simple, Portable, Scalable SMP Programming, http://www.openmp.org/.
[26] IBM, IBM synthetic data age code, http://www.almaden.ibm.com/software/mission/Resources/in dex.shtml.
[27] Bryan, G., T. Abel, and M. Norman, Achieving Extreme Resolution in Numerical Cosmology Using Adaptive Mesh Refinement: Resolving Primordial Star Formation, SuperComputing, Nov. 2001.
[28] B. Srinivas, Shoban Babu Sriramoju, "A Secured Image Transmission Technique Using Transformation Reversal" in “International Journal of Scientific Research in Science and Technology”, Volume-4, Issue-2, February-2018, 1388-1396 [Print ISSN: 2395-6011 | Online ISSN: 2395-602X]
[29] B. Srinivas, Gadde Ramesh, Shoban Babu Sriramoju, “A Study on Mining Top Utility Itemsets In A Single Phase” in “International Journal for Science and Advance Research in Technology (IJSART)”, Volume-4, Issue-2, February-2018, 1692-1697, [Online ISSN: 2395-1052]
[30] B. Srinivas, Gadde Ramesh, Shoban Babu Sriramoju, “An Overview of Classification Rule and Association Rule Mining” in “International Journal of Scientific Research in Computer Science, Engineering and Information Technology”, Volume-3, Issue-1, February-2018, 643-650 [ISSN : 2456-3307]
[31] B. Srinivas, Shoban Babu Sriramoju, “Managing Big Data Wiki Pages by Efficient Algorithms Implementing In Python” in “International Journal for Research in Applied Science & Engineering Technology (IJRASET)”, Volume-6, Issue-II, February-2018, 2493-2500, [ISSN : 2321-9653]
[32] Shoban Babu Sriramoju, “Analysis and Comparison of Anonymous Techniques for Privacy Preserving in Big Data” in “International Journal of Advanced Research in Computer and Communication Engineering”, Vol 6, Issue 12, December 2017, DOI 10.17148/IJARCCE.2017.61212 [ISSN(online) : 2278-1021, ISSN(print) : 2319-5940]
[33] Shoban Babu Sriramoju, " Review on Big Data and Mining Algorithm" in “International Journal for Research in Applied Science and Engineering Technology”, Volume-5, Issue-XI, November 2017, 1238-1243 [ISSN : 2321-9653], www.ijraset.com
[34] Shoban Babu Sriramoju, “OPPORTUNITIES AND SECURITY IMPLICATIONS OF BIG DATA MINING” in “International Journal of Research in Science and Engineering”, Vol 3, Issue 6, Nov-Dec 2017 [ISSN : 2394-8299].
[35] Shoban Babu Sriramoju, "Heat Diffusion Based Search for Experts on World Wide Web" in “International Journal of Science and Research”, https://www.ijsr.net/archive/v6i11/v6i11.php, Volume 6, Issue 11, November 2017, 632 - 635, #ijsrnet
[36] Dr. Shoban Babu Sriramoju, Prof. Mangesh Ingle, Prof. Ashish Mahalle “Trust and Iterative Filtering Approaches for Secure Data Collection in Wireless Sensor Networks” in “International Journal of Research in Science and Engineering”, Volum 3, Issue 4, July-August 2017 [ISSN : 2394-8299]
[37] Dr. Shoban Babu, Prof. Mangesh Ingle, Prof. Ashish Mahalle, “HLA Based solution for Packet Loss Detection in Mobile Ad Hoc Networks” in “International Journal of Research in Science and Engineering”, Volume 3, Issue 4, July-August 2017 [ISSN : 2394-8299]

[38] Shoban Babu Sriramoju, "A Framework for Keyword Based Query and Response System for Web Based Expert Search" in "International Journal of Science and Research" Index Copernicus Value(2015):78.96 [ISSN : 2319-7064].
[39] Sriramoju Ajay Babu, Dr. S. Shoban Babu, "Improving Quality of Content Based Image Retrieval with Graph Based Ranking" in "International Journal of Research and Applications", Volume 1, Issue 1, Jan-Mar 2014 [ISSN : 2349-0020]
[40] Dr. Shoban Babu Sriramoju, Ramesh Gadde, "A Ranking Model Framework for Multiple Vertical Search Domains" in "International Journal of Research and Applications" Vol 1, Issue 1,Jan-Mar 2014 [ISSN : 2349-0020].
[41] Mounika Reddy, Avula Deepak, Ekkati Kalyani Dharavath, Kranthi Gande, Shoban Sriramoju, "Risk-Aware Response Answer for Mitigating Painter Routing Attacks" in "International Journal of Information Technology and Management", Volume VI, Issue I, Feb 2014 [ISSN : 2249-4510]
[42] Mounica Doosetty, Keerthi Kodakandla, Ashok R, Shoban Babu Sriramoju, "Extensive Secure Cloud Storage System Supporting Privacy-Preserving Public Auditing" in "International Journal of Information Technology and Management", Volume VI, Issue I, Feb 2012 [ISSN : 2249-4510]
[43] Shoban Babu Sriramoju, "An Application for Annotating Web Search Results" in "International Journal of Innovative Research in Computer and Communication Engineering" Vol 2,Issue 3,March 2014 [ISSN(online) : 2320-9801, ISSN(print) : 2320-9798]
[44] Shoban Babu Sriramoju, "Multi View Point Measure for Achieving Highest Intra-Cluster Similarity" in "International Journal of Innovative Research in Computer and Communication Engineering" Vol 2,Issue 3,March 2014 [ISSN(online) : 2320-9801, ISSN(print) : 2320-9798]
[45] Shoban Babu Sriramoju, Madan Kumar Chandran, "UP-Growth Algorithms for Knowledge Discovery from Transactional Databases" in "International Journal of Advanced Research in Computer Science and Software Engineering", Vol 4, Issue 2, February 2014 [ISSN : 2277 128X]
[46] Shoban Babu Sriramoju, Azmera Chandu Naik, N.Samba Siva Rao, "Predicting The Misusability Of Data From Malicious Insiders" in "International Journal of Computer Engineering and Applications" Vol V, Issue II,Febrauary 2014 [ISSN : 2321-3469]
[47] Guguloth Vijaya, A. Devaki, Dr. Shoban Babu Sriramoju, "A Framework for Solving Identity Disclosure Problem in Collaborative Data Publishing" in "International Journal of Research and Applications", Volume 2, Issue 6, 292-295, Apr-Jun 2016 [ISSN : 2349-0020]
[48] Monelli Ayyavaraiah, Shoban Babu Sriramoju, "A Survey on the Approaches in Targeting Frequent Sub Graphs Mining" in "Indian Journal of Computer Science and Engineering (IJCSE)", Volume 9, Issue 2, Apr-May 2018 [e-ISSN : 0976-5166 p-ISSN : 2231-3850], DOI : 10.21817/indjcse/2018/v9i2/180902024
[49] Monelli Ayyavaraiah, "Review of Machine Learning based Sentiment Analysis on Social Web Data" in "International Journal of Innovative Research in Computer and Communication Engineering" Vol 4,Issue 6,March 2016 [ISSN(online) : 2320-9801, ISSN(print) : 2320-9798]
[50] Monelli Ayyavaraiah, " A Study on Large-Scale Cross-Media Retrieval of Wikipedia Images towards Visual Query and Textual Expansion" in "International Journal for Research in Applied Science and Engineering Technology", Volume-6, Issue-II, February 2018, 1238-1243 [ISSN : 2321-9653], www.ijraset.com
[51] Monelli Ayyavaraiah, "Nomenclature of Opinion Miningand Related Benchmarking Tools" in "International Journal of Scientific & Engineering Research" Vol 7,Issue 8, February 2018, [ISSN 2229-5518]
[52] Siripuri Kiran, 'Decision Tree Analysis Tool with the Design Approach of Probability Density Function towards Uncertain Data Classification', International Journal of Scientific Research in Science and Technology(IJSRST), Print ISSN : 2395-6011, Online ISSN : 2395-602X,Volume 4 Issue 2, pp.829-831, January-February 2018. URL : http://ijsrst.com/IJSRST1841198
[53] Ajmera Rajesh, Siripuri Kiran, " Anomaly Detection Using Data Mining Techniques in Social Networking" in "International Journal for Research in Applied Science and Engineering Technology", Volume-6, Issue-II, February 2018, 1268-1272 [ISSN : 2321-9653], www.ijraset.com