



## Data Mining and Optimization: A Survey Paper

Sushma Lata

M.Tech. Student, NRI College,  
Bhopal, India

**Abstract:** This paper summarizes the data mining concept and its implications in the direction of its approaches. The data mining based on optimization using iterative particle swarm optimization(PSO) is researched in detail and the association rule mining know-how and ways is used to obtain relationships among item sets. The paper also conducts a proper assessment of the expanse of frequent pattern mining algorithms and systolic tree architecture which is characterized by high throughput and faster execution.

**Keywords:** Data Mining, Web Mining, PSO, web worth, frequent pattern mining, systolic tree, random velocity, Association Rules.

### I. INTRODUCTION

Data mining refers to extracting enormous amount of data from a large data set and transforming that to understandable form. Association rules in distributed data mining is also an area of data mining which finds association rules over items physically throughout the network. In case of e-commerce sites, a standard client-server model is followed having three components: the server system, the network and the client system and protection of each is necessary from any hazard and occurrence. Web mining is related with extracting knowledge from web data and is categorized as: **Web usage mining, Web content mining** and **Web structure mining**. Frequent pattern mining is related to extracting the patterns which occur frequently, a pattern can be a form, a template or a set of rules or a data behaviour.

Distributed data mining too is related with extracting very intricate set of relations among entities. Distributed data mining is used to extract extremely hidden data that is very difficult to find out. Distributed data mining has off late gained a lot more importance due to the expansion of Internet and its wider reach. Association rules are applied as extraction rules to discover interesting set of relations among a group of objects or items.[1]

### II. SYSTOLIC TREE

A systolic tree is an organization of pipelined processing elements (PEs) in a multidimensional tree pattern.[2] It is not constantly useful or effectual to directly interpret a software procedure into a hardware design. When the definite node degree at specific point in the tree overdoes the projected node degree, certain recurrent itemset may not be traced.[3]

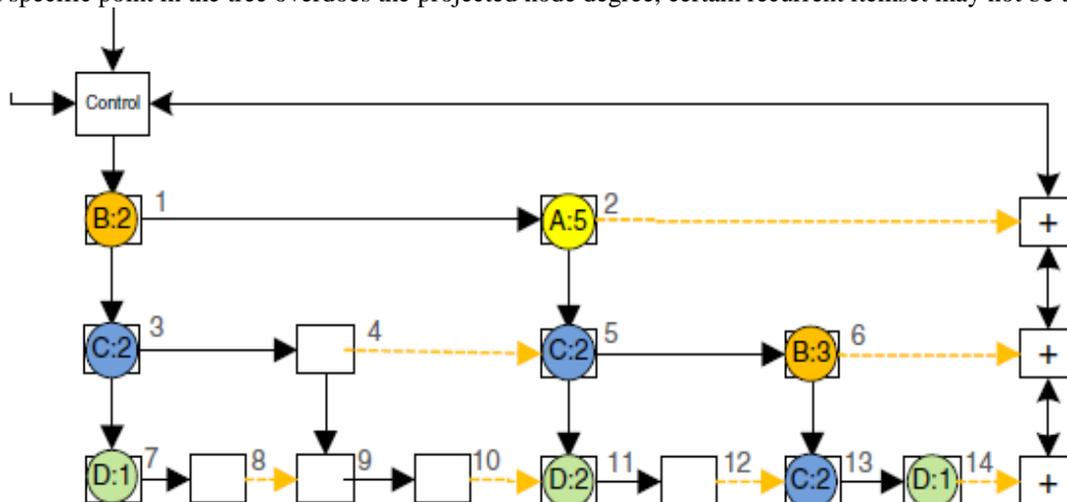


Fig. 2 The systolic tree architecture

Every PE (Processing Element) takes single input from the parent besides two outputs to its child and siblings. Every PE only partakes a linking by its left child. If it takes to direct the data to its right child, the data is delivered to the left child then through siblings, transient over all offspring on the pathway. The overall processing elements which hold no item are unfilled. If the objects in single operation are relocated into the design in a rising order, some PE must comprise a minor element instead of that of its children. Though this ensures no surity that the node element in the systolic tree is continuously superior than its left-side siblings.[3]

### III. WEB WORTH OPTIMIZATION

Optimizing a website can refer to search engine optimization, which surges a website's ranking on search engines and ensuring that website runs speedily and effortlessly. Diffusion of knowledge is the apparatus that pushes scientific progress. Because improvements hinge principally on prior observations.

### IV. PSO (PARTICLE SWARM OPTIMIZATION)

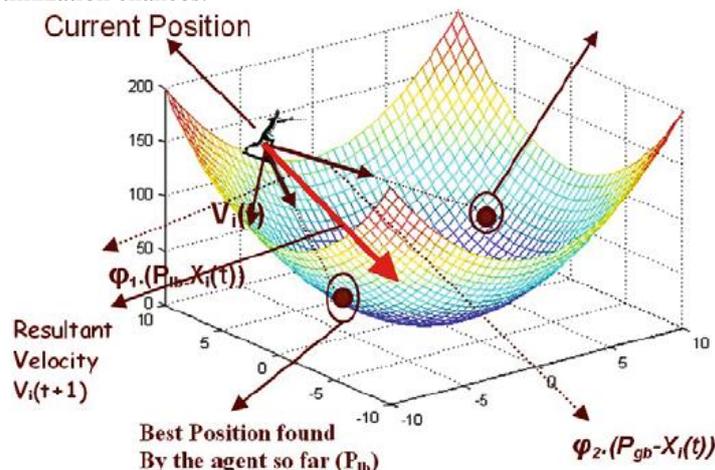
In computer discipline, **particle swarm optimization (PSO)** is a computational technique that improves a difficulty by iteratively trying to recover an entrant solution with respect to a given degree of value. PSO optimizes a problem by having a number of candidate results, also called particles, and moving these particles everywhere in the search-space conferring to unassuming mathematical formulae over the particle's position and velocity. Each particle's movement is subjective to its native best recognized position, nonetheless, is also directed to the finest known locations in the search-space, which are restructured as improved positions are established by other particles. This is anticipated to transfer the swarm en route for the best solutions and is an algorithm for evolutionary optimization[4].

Being a multi-agent parallel search technique, particles are deliberated as conceptual entities, flying over the multi-dimensional pursuit space. At any specific

instant, every particle has a point and a velocity. The locus vector of a particle with reverence to the source of the search space symbolizes a trial solution of the search difficulty. In the starting, a population of particles is initialized with random positions. The population of such particles is called a "swarm"  $S$ . A neighborhood relation defines whether two particles are neighbours or not. This neighbourhood topology affects the swarm performance. Each particle has two dimensions current position and current velocity and a small memory recorded with its previous position[10]. When the particles are initialized, an iterative optimization starts as a process and position and velocity are altered by mathematical functions. The iteration continues until a suitable solution has been found or CPU usage has reached an upper limit.

Parameters are selected such that they optimize the search space but they behave differently in different situations. Therefore we must know their probable effect on the solutions and thus must be chosen carefully. Controlling the particle momentum is essential so an inertia weight is chosen. Velocity of the particle depends upon that. Throughout iteration particle goes on change in its particular position which is determined by maximum velocity. Constriction coefficient results in converging particles over time. The particles oscillate and their collapse is avoided by this coefficient. When the previous and present best conditions are far apart, particle does a wide search and based upon some previous results the final result is predict.

Thereby increasing the optimization chances.



The above figure demonstrates how PSO finds the best possible solution .

But it should be borne in the mind that no of particles should be limited to the range 20-60 ,the greater is the swarm size the higher is the probability of convergence of errors.PSO can be either global or local .Global relate to some high standard of frequency and local is limited to some limited range of frequency.

In the former the velocity is adjusted dynamically while in the latter it depends upon the searches within the neighborhood .

### V. WEB MINING AND E-COMMERCE

Web usage has grown in leaps and bounds,therefore data extraction too has become cumbersome task.It is important in order to improve the site and enhance the chance of fetching advertisements.Classification of data becomes necessary and association rules are applied to relate data items and their relations[3]

### VI. CONCLUSIONS

Data mining is in nascent stage of its research.It scope is getting wider day by day .Data is increasing on web and for the users to access the data has become quite difficult.Here optimization plays its role in helping the users to get the optimal data.Data mining also helps in automated discovery of previously unknown information which make it to have a wider reach in the Information technology field.

**REFERENCES**

- [1] Mining Rare Association Rules in a Distributed Environment using Multiple Minimum Supports :Jutamas Tempaibookul,School of Engineering and Technology Asian Institute of Technology Pathumthani, Thailand
- [2] An Efficient Weighted Rule Mining for Web Logs Using Systolic Tree P. Sampath, C. Ramesh, T. Kalaiyarasi, S. Sumaiya Banu, G. Arul Selvan<sup>2</sup> Department of Computer Science & Engineering, Bannari Amman Institute of Technology Sathyamangalam, Tamil Nadu, India Department of Computer Science & Engineering, EGS Pillay Engineering College, Nagapattinam, Tamil Nadu, India “IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM -2012) March 30, 31, 2012 432
- [3] Mahendra Pratap Yadav , Mhd Feeroz, Vinod Kumar Yadav” Mining the customer behavior using web usage mining in e-commerce” IEEE-20150
- [4] Swagatam Das, Ajith Abraham, and Amit Konar” Particle Swarm Optimization and Differential Evolution Algorithms: Technical Analysis,Applications and Hybridization Perspectives”