

A Novel hybrid comparison based on Simulated Annealing techniques

Sowmya k Menon¹ and Varghese Paul²

¹ Associate Professor, Department of Computer Applications, SNGIST ,
N.Paravur, Cochin, Kerala, India

² Professor in IT, Department of IT , Rajagiri school of engineering and Technology,
Kalamassery, Cochin, Kerala, India

Abstract

As the availability of data is increasing day by day, clustering is one of the fastest growing research areas. There are many clustering techniques. K-means is one of the simplest clustering techniques. In this paper we are surveying the usage of simulated annealing in different fields. Simulated annealing is used to solve the problem of local minima. This function searches for global minimum of a very complex non linear objective function with a very large number of optima. Simulated annealing can be used to locate the cluster centres. Simulated annealing mechanism is used to locate the cluster centres. This is mainly used as a part of code optimization of k-means. One advantage of k-means is that it is used to cluster large datasets and its performance increases as large number of datasets increases.

Keywords: *Simulated Annealing, k-means, clustering, time tabling, Hyper heuristics, Meta heuristics, Hybrid approach*

1. Introduction

Clustering is found to be a basic process to human understanding. The grouping of objects based on clustering is used in relevant areas such as statistics, economics, physics, psychology, biology, engineering, marketing etc. Simulated annealing methodology can be used in timetabling problem for allocating different timeslots for lecturers. There may be hard constraints and soft constraints. For getting feasible solution don't violate the hard constraints.

M.Nandhini and Dr.S.Kanmani[1](2009) defined timetabling as a problem and can be solved by simulated annealing techniques. SA is similar to artificial intelligence. This problems in time table slots in schools and universities can be solved using SA, artificial intelligence etc. When the problems have larger size and if we want to find an optimal

solution we are concentrating on SA techniques. The timetabling problem can be solved using, heuristic, meta heuristic and hybrid approaches. Since+ Annealing has a great role in metallurgy. This technique can be used in controlling heating and cooling of a material. The clustering problem domain can be illustrated using two domain models 1. A data structure used to define clusters 2. An internal clustering criterion based on model of clusters. Since it is a real world problem and depends on the institution where it is applied, constraints and objective function varying accordingly. Even then, finding standard framework/solution is difficult task. Number of recent researches have been going on with heuristics, meta heuristics and hybridization The optimization of the problem depends on response time, work of the solution and space occupied. In order to attain optimal solution ,constraints, heuristic selection and algorithmic selection are very important.

Donald E Brown and Christopher L Huntley[2](july1991) formulated a practical application of simulated annealing to clustering. Careful selection of clustering criterion is very much necessary. We use simulated annealing to find out the nearest optimal clusterings for each set of criteria. By comparing these optimal clusterings with the true clusterings using an external clustering criterion.

Stephen Merendino and M .Emre Celebi[3] focuses on simulated annealing clustering algorithm based on center perturbation using Gaussian mutation. This paper focuses on giving a detail on how to improve the clarity of clustering. Clustering is grouping of similar data sets. The clustering problems can be decreased by using certain function and simulated annealing algorithms have been developed to solve a variety of optimization problems. SA is always based on physical annealing process. Metropolis algorithm

has been developed to this physical annealing process. In physical annealing process as an example when a material is heated in a heating bath and allows the material to cool and changes the material to low energy state.

S. Kirkpatrick, C.D.gelatt,Jr.M.P.Vecchi[4](1983) focus on optimization by aimulated annealing. This article mainly focuses on optimization and statistical mechanics and develop similarities between two fields.

Abha Kaushiki,Suhajit Ghosh and Sunita Kumari[5](2014) paper focuses on improving the k-means clustering process using simulated annealing techniques. In this paper inorder to improve the quality of clustering(SSE) is used ie, Sum of squared errors.

2. Simulated annealing

Simluated Annealing is used to solve the problem of local minima. This function searches for global minimum of a very complex non-linear objective function with a very large number of optima. Simulated annealing techniques are used in problem solving approaches such as heuristics, meta heuristics and hybrid approaches. Many optimization problems can be solved using simulated annealing techniques.

[5] SA has a great role in searching mechanism. This is mainly used for the purpose of determining a fixed number of cluster centres. Annealing mechanism has been used in controlling the heating and cooling of material, thereby increasing the size of crystals and therefore help to reduce the defects.

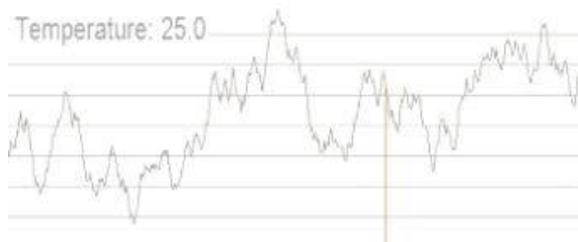


Fig1: Temperature variation of a material due to SA

Simulated annealing can be used to find the global minimum of a cost function that possess certain local minimum.

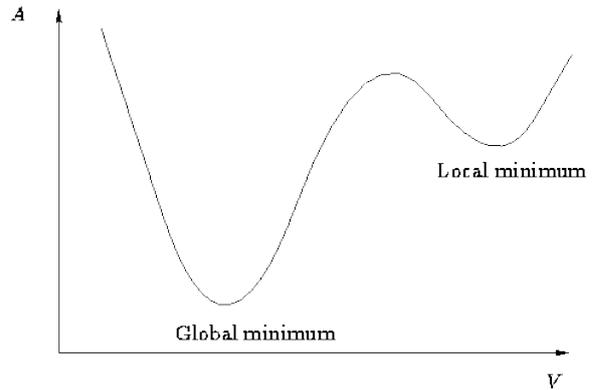


Fig 2: Local and global minimum

Simulated annealing can be used in statistical applications. Many researches think that SA is a tool for getting optimal solution. Simulated annealing can be sought out as a powerful optimization technique used to find global minimum that can be derived from statistical mechanics. SA is used for comparison purposes. Simulated annealing is applicable to partitional clustering but the implementation can't be easily done using hierarchical clustering.

3. Comparison of Research works

[1] explains about the timetable problem and how it can be solved using simulated annealing techniques. Many problem solving approaches such as heuristics, meta heuristics and hybridization can be used in this. The time tabling optimization mainly depends on the institution in which it is applied. This paper is surveying the simulated annealing techniques and the usage of heuristics and meta heuristics approaches for problem solving in time tabling. [1] explains SA as a local search technique. Here the three phases of hybrid approach is also explained.

[2] presented a new SA based partitional clustering algorithm called SAGM. This SAGM algorithm is very fast and efficient compared to other SA algorithms. The SAGM algorithm is not swapping the data points between clusters, instead it uses a Gaussian mutation to perturb a randomly chosen cluster center in a controlled manner.

[3] In this paper explains about SINICC, an empirical method for comparing the internal clustering criterion. Simulated clustering also provides a best clustering criterion. SINICC is mainly using simulated annealing techniques. [3] also explains pattern recognition. In pattern recognition, clustering can be used to find patterns in data without supervising it. This paper focuses on formulating clustering as an optimization problem using a used defined objective function called internal clustering criterion. [3] found out a method called SINICC which uses simulated annealing to

find the optimal solution to the partitional clustering problem.

The paper focuses on an application of SINICC, in which the data sets in this are inexact environment objects. The performance of SINICC has applications in other problems in pattern recognition. This work gives a clear evidence that simulated annealing is used for evaluation of partitional clustering methods. Previous work in this paper proved that simulated annealing is impractical and introduced SINICC in simulated annealing.

[4] This paper shows how metropolis algorithm can be used for simulation of body systems at a finite temperature. This idea is being applied this to a number of problems arising in optimal design of computers. One example of optimization problem is traveling salesman that requires more intensive study. This paper tells about detailed explanation of statistical mechanics and optimization. Implementing the metropolis algorithm in simulated annealing and combinatorial optimization problem can be made effective in future.

[5] focuses on optimization approaches to k-means clustering using simulated annealing. In this simulated annealing shows the results of k-means clustering. In this paper many examples are pointed out in which simulated annealing acts as hill climbing. This paper focuses on error estimation of k-means. By evaluating and comparing two different aspects of k-means we can evaluate its quality. The error can be estimated using the formulae of sum of squared error ie, SSE.

$$SSE = \sum_{i=1}^K \sum_{x \in c_i} d(x, c_i)^2$$

Table1. Table showing research works

Yearr	Article name	Author	Approaches	Algorithm	Methods used
2009	Survey on Simulated annealing methodology for university course time tabling	M.Nandhini, Dr.S.kanmani	Hyper heuristics, hybrid approaches, meta heuristics	GGA algorithm	Simulated annealing
2014	Survey on optimization approaches to k-means using simulated annealing	Abha Kaushik,Subhajit Ghosh & Sunitha kumari	Optimization	k-means algorithm	SA
2013	Simulated annealing clustering algorithm based on center perturbation using Gaussian mutation	Stephen merendino, Emre Celebi	Gaussian mutation	Metropolis algorithm, Conventional SA algorithm	SA
1991	A practical application of simulated annealing to clustering	Donald E Brown & Christopher L Huntley	Simulation of near optima for internal clustering criteria	Metropolis algorithm	SA
1983	Optimization based on simulated annealing	S.Kirkpatrick, C.D.Gelatt, Jr. M.P.Vecchi	Optimization	-	SA

Simulated annealing is a searching optimization based on Monte Carlo iteration strategy. This paper conclude the advantage of K-means and explains that the performance of k-means can be increased by increasing the number of clusters. The paper first tells about the basics of data mining and explores various research papers related to this. Many research works are being carried out in simulated annealing and K-means clustering.

Simulated annealing is an optimization techniques and the implications of simulated annealing are used in different research works. Researches are also being carried out in this area. Section must contain specific details about the materials studied, instruments used, specialized chemicals source and related experimental details which allow other research worker to reproduce the results. The journal will not be held responsible if any kind of plagiarism followed and the editor's decision would be final if any litigation arises during processing or after publishing.

4. Proposed work

The proposed work is an implementation of a novel recommendation engine using simulated annealing based clustering. Till now the analysis ,design, testing has been completed. In the future, planning to optimize the k-means code for simulated annealing based clustering for the web recommendation system.,

5. Conclusion

The paper is a survey of the simulated annealing techniques. This paper focuses on the contributions of simulated annealing techniques in various aspects. In my work, optimization of cluster algorithm was done using simulated annealing techniques. Simulated annealing helps to overcome the local minima problem by choosing the best cluster number

References

- [1]. Jia Li and Osmar R. Ziane, "Combining Usage, Content, and Structure Data to Improve Web Site Recommendation", *In Proceedings of the 5th International Conference, EC-Web*, Spain, 2004, pp 305-315. Reference 1
- [2]. Harita Mehta, Shveta Kundra Bhatia, Punam Bedi and V. S. Dixit, (November 2011). Collaborative Personalized Web Recommender System using Entropy based Similarity Measure. *IJCSI International Journal of Computer Science Issues*, 8(6), pp. 231-240.
- [3]. Toby Segaran, *Programming Collective Intelligence*, 1st ed, O'Reilly, 2007.
- [4]. G. Adomavicius and A. Tuzhilin, (2005). Towards the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Trans. on Data and Knowledge Engineering*, 17(6), pp. 734-749.
- [5]. I.J Witten, E. Frank, "Data Mining: Practical Machine Learning Tools and Techniques," Elsevier, 2011.