**Robust Scalable Visualized Clustering in Vector and non Vector**

**Semimetric Spaces, G Fox - Parallel Processing Letters,**

**2013 - World Scientific**

**The Summary**

Mashar Cenk Gencal

School of Informatics and Computing,

Indiana University, Bloomington

mcgencal@indiana.edu

**Abstract**

 In this paper, a new idea is revealed to data analytics for large systems by using robust parallel algorithms which runs on clouds and HPC systems. However, it is assumed that the data is already described in a vector space, and the pairwise distances between points are also described. Moreover, it is achieved to make some known algorithms better for functionality, features and performance. Furthermore, it is mentioned about steering complex analytics where visualization is important for both the non vector semimetric case and clustering high dimension vector spaces. Deterministic Annealing is used to have fairly fast robust algorithm. The method is practiced in some life sciences applications.

**1 Introduction**

 In the publication, deterministic annealing is defined, and then it is considered that clustering with an highlight on some of features which are not found in the openly available software such as R. After that, it is mentioned about difficulty of parallelization for heteregeneous geometries. Finally, in the last section, it is suggested to use a data mining algorithm with some form of visulization since the most of datasets do not show that whether or not an analysis is adequante. Because of that, it is advised that “ The well studied area of dimension reduction deserves more systematic use and show how one can map high dimension vector and non vector spaces into 3 dimensions for this purpose.” [1]

**2 Deterministic Annealing**

 Before explaining the idea in the publication, we need to explain Determistic Annealing, Gibbs Distribution, Free Energy F :

**2.1 Gibbs Distribution, Free Energy**

Gibbs Measure: (1)

where

 ,

 ,

 ,

 (2)

* if , each point is equally associated with all clusters.
* if , each point belongs to exactly one cluster

Total partition function : (3)

Expected energy of x;

 (4)

Thus, total energy is,

 (5)

If F is effective cost function ( free energy ), from the equation (3) ;

 (6)

Therefore, if , then

 where is the representative of (7)

Thus, the probability that x belongs to from equations (1) , (2) and (7) ;

 (8)

From equations (6) and (8) ;

 (9)

If we take derivation of this function, we will have ;

Hence, if we keep taking derivatives of the function, from (9), we will have ;

 ,

where f is based on , (10)

**2.2 Deterministic Annealing Algorithm**

1. Set
2. Choose an arbitrary set of initial cluster means
3. Iterate according to equation (10) until converged
4. Increase
5. If go to step 3.

**2.3 The Idea in The Publication**

 Optimization problems are considered by finding global minima of energy functions in the publication. Since Monte Carlo approach of simulated annealing is too slow, integrals are analytically applied by utilizing some different approximations in statistical physics. Hamiltonian, H(x), is minimized for the variable, χ . Then, the idea relating to average the Gibbs Distribution at Temperature T is introduced.

P(χ) will denote probability distribution for χ of type T,

 (11)

or (12)

and S(P) will denote entropy associated with probability distribution P(χ),

S(P) = - P(χ) lnP(χ) (13)

After minimizing the Free Energy F by using Objective Function and Entropy,

 (14)

 For each iteration, the temperature is decreased from 0.95 to 0.9995. Since it has difficulty for some cases, e.g. vector space clustering, Mixture Model, the new Hamiltonian, H0(χ, ε), is revealed.

 (15)

F() = < H0 - T S0(P0) >|0 = < H– H0> |0 + F0(P0) (16)

where < ….. >|0 shows that averaging with Gibbs i.e. . ( is fixed in these integrals)

Since P minimizes F(P), we will have ;

 (17)

 This helps us for EM method, which is to calculate the expected value of the log likelihood function, and to find the parameter that maximizing the quantity. Therefore,

<χ> = <χ> |0 = ∫ dχ χ Po(χ) (18)

 For the new Hamiltonian function, there are three variables: It is descriped that “ The first variable set are used to approximate the real Hamiltonian H() by H0(ε , χ, ) ; the second variable set x are subject to annealing while one can follow the determination of x by finding yet other parameters optimized by traditional method.” [1]

 It is said that the deterministic idea has some difficulties such show to choose parameter. Also, it is challenging to minimize the free energy from the equation (16).

**3 Vector Space Clustering**

**3.1 Basic Central or Vector Space Clustering**

 It is assumed that we have a clustering problem with N points labeled by x and K clusters labeled by k. If is the probability that point x belongs to cluster k with constarint,

 (19)

where (20)

 : points’ position

 : cluster center

Clearly, we can write the equation (19) as follows :

 (21)

If we rewrite the equations (15) and (16) ;

 (22)

F() (23)

After applying the equation (21) and (23) to the equation (22) ;

 (24)

The optimal vectors are derived by maximizing the entropy of the Gibbs distribution.[5]

 (25)

If we take derivative of the equation (25), and set the result 0 ;

 (26)

And the cluster centers Y(k) are easily determined in M step.[1]

 (27)

The non vector case is not mentioned in detail. However, we have the equation ;

 (28)

where is the number of points in k th cluster

 and is the total number of points

**3.2 Fuzzy Clustering and K-means**

 If the temperature is very high, the exponent of the equation (26) goes to zero, and all centers have equal probability. In that case, all centers have the position,

 (29)

 If the temperature decreases, the points tend to close to a center by stages, and in the zero temperature limit, K-means solution can be found by assigning each point with the nearest center.



 Figure 1 declares the idea behind the annealing which is that we could not find false minima at high (infinite) temperature where the object function is smooth, and the free energy is minimized. If we decrease the temperature slowly, we close to the minima at lower temperature. Hence, the annealing is robust as long as temperature lowered slow enough. [1] In the figure, temperatures T2 and T3 show that the false minima at finite temperature. Tihs process is being approximated in deterministic annealing. The approximation includes modifying integrals by evaulating functions through their mean values.

**3.3 Multiscale ( Hierarchical ) and Continous Clustering**