International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:10, No:11, 2016

Chemical Reaction Algorithm for Expectation Maximization Clustering

Li Ni, Pen ManMan, Li KenLi

***Abstract*—**Clustering is an intensive research for some years because of its multifaceted applications, such as biology, information retrieval, medicine, business and so on. The expectation maximization (EM) is a kind of algorithm framework in clustering methods, one of the ten algorithms of machine learning. Traditionally, optimization of objective function has been the standard approach in EM. Hence, research has investigated the utility of evolutionary computing and related techniques in the regard. Chemical Reaction Optimization (CRO) is a recently established method. So the property embedded in CRO is used to solve optimization problems. This paper presents an algorithm framework (EM-CRO) with modified CRO operators based on EM cluster problems. The hybrid algorithm is mainly to solve the problem of initial value sensitivity of the objective function optimization clustering algorithm. Our experiments mainly take the EM classic algorithm:k-means and fuzzy k-means as an example, through the CRO algorithm to optimize its initial value, get K-means-CRO and FKM-CRO algorithm. The experimental results of them show that there is improved efficiency for solving objective function optimization clustering problems.

***Keywords*—**Chemical reaction optimization, expectation maximization, initial, objective function clustering.

I. Introduction

CLUSTERING is a very important non-supervised data analysis method. It is the processing of dividing the data objects into multiple clusters, which makes the objects in the cluster have high similarity, but the similarity between clusters is not high [9], [11], [18]. Clustering as a data mining tool has been referenced to a lot of fields, such as biology, security, big data processing [4], [15], etc.

The expectation maximization (EM) [19] is a kind of algorithm framework in clustering method, based on the initial parameter set, and iterated until the convergence of the clustering algorithm, for example, K-means algorithm and fuzzy c-means algorithm. Each iteration contains two steps: Expectation step (E-step): The center of a given cluster, each pattern is assigned to the clustering which include the nearest cluster center from the pattern; Maximization step (M-step): Adjust the center for each cluster, then the distance sum between clustering pattern and new center is minimized, the similarity in clusters is maximized. So EM algorithm is a kind of optimization algorithm based on the objective function. In order to reach the global optimum, optimization of cluster objection function may exhaust all possible, the amount of calculation is very great, and has also been demonstrated that in the general Euclidean space, even only two clusters, the problem is NP hard problem. Therefore, in the practice, it tends to use the heuristic method, gradually improve the clustering quality, approach global optimal solution.

In order to optimize the clustering algorithm, a lot of evolutionary algorithms are introduced into the clustering problem [2], [3], [6], [7], [16]. Such as simulated annealing algorithm, genetic algorithm, etc. It knows that the evolutionary algorithm is random search, and the initial value is not sensitive to the algorithm. However, the EM algorithm is sensitive to the initial value, to achieve the global optimal value. So, it can use the latest optimization technology to achieve its optimization process.

In this paper, the chemical reaction algorithm (CRO) [1] is combined with the EM algorithm to achieve the global optimization clustering algorithm.

II. EM Algorithm

EM algorithm [19] is a method for the estimation of parameters of the maximum likelihood estimation which is proposed by Dempster, Laind, Rubin in 1977, is one of the ten algorithms of machine learning, often used in machine learning and data clustering field. EM algorithm processing is as follows:

The first step: Initialization distribution parameters

The second step: Repeat until convergence

E step: To calculate the expected value of unknown parameters by using the existing estimation of distribution parameters, and get the current parameter estimation.

M step: Repeated estimation of distribution parameters, making the maximum likelihood of the data.

The main goal of EM algorithm is to provide a simple iterative algorithm to optimize the objective function, which is the biggest advantage is simple and stable, but it is easy to fall into local optimum.

Based on the prototype of the clustering analysis technology, the most important use of EM algorithm is: K-means clustering and fuzzy K-means clustering. In this paper, the optimization problem of EM algorithm is studied for the two types of EM algorithm as examples.

A. Fuzzy K-Means (FKM) Algorithm

With the introduction of the concept of Fuzzy Division, Bezdek gives a general description of the fuzzy clustering based on the objective function [8]–[12], [14], [17], and the objective function becomes the weighted sum of squared errors

L. Ni was with College of Computer Science and Electric Engineering, Hunan University, Hunan, Changsha 410082 China and Department of Information Science and Engineering, Hunan City College, Hunan, Yiyang (e-mail: lini66600079@163.com).

P. ManMan and L. KenLi are with Hunan University.

(SSE). If it has known data set X, the number of clustering K, the weight of m,the objective function is expressed as:

$$J_m(U,C) = \sum_{i=1}^{i=n}\sum_{c=1}^{c=k}(u_{ci})^m(d_{ci})^2 \quad (1)$$

and condition as: $\sum_{c=1}^{c=k} u_{ci} = 1$. The criterion of clustering is to calculate the minimum value of $J_m(U,C)$. They have used the Lagrange multiplier method to solve them:

$$F = \sum_{c=1}^{c=k}(u_{ic})^m(d_{ic})^2 + \lambda(\sum_{c=1}^{c=k} u_{ci} - 1) \quad (2)$$

A necessary condition for the first order of the optimization problem:

$$\frac{\partial F}{\partial \lambda} = \sum_{c=1}^{c=k} u_{ci} - 1 = 0 \quad (3)$$

$$\frac{\partial F}{\partial u_{jt}} = m(u_{jt})^{m-1}(d_{jt})^2 - \lambda = 0 \quad (4)$$

$$\frac{\partial F}{\partial p_c} = \sum_{i=1}^{i=n}(u_{ci})^m(x_i - p_c) = 0 \quad (5)$$

so, FKM proceeds by iterating through the equation for computing the membership values u_{ci},and cluster prototype P_k, as:

$$u_{ci} = \frac{1}{\sum_{l=1}^{l=k}(\frac{d_{ci}}{d_{li}})^{\frac{2}{m-1}}} \quad (6)$$

$$p_c = \frac{\sum_{i=1}^{i=n}(u_{ci})^m x_k}{\sum_{i=1}^{i=n}(u_{ci})^m} \quad (7)$$

B. K-Means Algorithm

K-means algorithm [20] is the most widely used data clustering method. K-means algorithm is the input of the number of clusters K, and n data objects of the database, the output to meet the minimum variance of the standard k clustering. It used the "proximity metric" to quantify the "nearest" concept, and It usually uses "distance" as a measure of similarity. The prototype of the cluster is the best representative of the cluster point.

The distance between samples is Euclidean distance, the error for each sample is Euclidean distance between it and its nearest prototype, their sum of error square (SSE) [20] is used as the objective function, the process using the EM method is k-means algorithms.

In multidimensional data, the Euclidean distance between the sample and the cluster prototype [20] is expressed as: $d_{ik}^2 = ||x_k - c_i||$. k Cluster prototype $c_k = \frac{1}{m_k}\sum_{x\in C_k} x_k$. Objective function: $SSE = \sum_{i=1}^{i=k}\sum_{x_k\in C_i} d_{ik}^2$.

So, it can be seen that K-means [20] is a special kind of FKM algorithm. The membership value u_ci of the sample X_i belongs to c cluster must be 0 or 1.

Note that, the EM method [19] has the advantages of simple and fast convergence speed, but because the algorithm is starting from an arbitrary initial values, the algorithm is sensitive for the initial value and different initial worth will get different clustering results.

C. Chemical Reaction Optimization (CRO)

CRO [1] is recently established metaheuristics for optimization, inspired by nature of chemical reactions, developed by Albert Lam and Li (2013) [1]. This algorithm is comparable with the genetic algorithm, ant colony algorithm and other intelligent optimization algorithms. In CRO, a candidate solution for a specific problem is encoded as a molecule. Each molecule has several essential properties: Molecular structure (ω), potential energy (PE), kinetic energy (KE) and so on. In the process of chemical reaction, the reactant always tends to be transformed into a more stable product with the minimal PE. In essence, CRO is a stochastic optimization method that enables the system achieve the minimum PE.

In the optimization problem, PE is defined as the objective function of the corresponding solution [1] represented by ω.

$$PE(\omega) = f(\omega) \quad (8)$$

The chemical reaction is caused by the collision. The basic introduction of algorithm is as follows: The four elementary reactions: Single molecule collision, decomposition, molecular collision and synthesis. Seven basic CRO parameters: popsize(), MoleColl, KELossRate, InitialKE, buffer, α, β. And four operators: Two-exchange, Gaussian perturbation with reflection, half-total change. Two laws of conservation of energy:

$$\sum_{k=1}^{k=popsize(t)} PE_k(t) + KE_k(t) + buffer(t) == C \quad (9)$$

$$\sum_{k=1}^{k=c} PE_k + KE_k \geq \sum_{k=1}^{k=l} PE'_k \quad (10)$$

The execution of CRO is as:

1) begin
2) initialization
3) judge rand()>MoleColl
4) 3)satisfied, then judge $KE \leq \beta$
5) 4)satisfied, synthesis
6) inter-molecular ineffective collisions
7) 4) didn't satisfied, synthesis
8) 3) didn't satisfied, molecule selection
9) judge $NumHit - MinHit > \alpha$
10) 10) satisfied, on-wall ineffective
11) decomposition
12) check for min PE
13) curFE<parameFElimit satisfied, return 3)
14) end

Overall, CRO is a algorithm framework based on the objective functions optimization problem. It can define and adjust molecular structure and energy management according to the user needs to solve practical problems. So CRO has a strong flexibility. In recent years, CRO algorithm has

successfully solved many complex problems [5], [6], [13], and the optimal solution is better than a lot of intelligent algorithms.

III. EM-CRO Algorithm

This section presents the global mechanism of EM-CRO starting from EM algorithms and CRO.

A. Molecule Reaction in n Dimension Space

1) Molecular Structure Representation: A molecular structure represents a complete solution. In EM clustering problem,the structure of a molecule represents the membership function. Then the structure of a molecule can be represented as:

$$M_i = [U(i,1)(t), U(i,2)(t), ..., U(i,n)(t)], 0 \leq i \leq c \quad (11)$$

It is the ith molecule of the population of molecules

2) Derivation of Potential Energy Function: In EM approach, this is to minimize the SSE function. Then the EM-CRO combines the clustering objective function and the characteristics of CRO algorithm. Set $potential\ energy = SSE$ as the objective function. Because it meets the requirements of PE reduction of CRO and clustering requirements with minimal error.

3) CRO Molecule Operation of Multidimensional Space: In the EM-CRO clustering method, solutions have to be optimized in n-dimensional search space according to the U molecule. So it is implemented the on-wall ineffective collision, decomposition, inter-molecular ineffective collision and synthesis operation in multidimensional space.

1) Molecular wall operation: It is a small change in the molecular structure of a single molecule in the independent space. Under the condition of conservation of energy the equivalent of the target function in any adjacent area search is $PE(\omega') = f(\omega')$. For example: out(i,j)=in(i,j)+randn.
2) Decomposition operation: The decomposition refers to the case when a molecule encounters a collision and then decomposes into two molecules. It is satisfied with $f(\omega) \geq f(\omega_1') + f(\omega_2')$.
3) Inter-molecular ineffective collision operation: It takes place when two molecules collide with each other, then dislocate the result, i.e. $out_1(i,j) = in_2(i,j), out_2(i,j) = in_1(i,j)(out(i,j)$ is 2-dimensional output molecule, in(i,j) is 2-dimensional input molecule).
4) Synthesis operation: It does the opposite of decomposition. In n-dimensional space, it merges two molecules into one molecule using $out(i,j) = in_1(i,j)$ or $out(i,j) = in_2(i,j)$.

B. Em-CRO Algorithm Framework

The basic steps of the EM-CRO algorithm are:

1) begin
2) input data set, k (cluster center number)
3) Initialization optimization using CRO: implement CRO optimization process based on objective function $J_m(U,C)$
4) compute the U based on (6), replace the existing U
5) compute the P based on (7), replace the existing cluster centers
6) compute $J_m(U,C)$
7) stop if termination criterion is attained, otherwise repeat from step 3)
8) end

TABLE I
SSE Value Obtained by FKM,FKM-CRO, K-Means and K-Means-CRO Clustering Algorithms for Artificial Data Sets: Data-60-2 and Data-2000-2

SSE	data-60-2	data-2000-2
FKM	169.9916	788.3131
FKM-CRO	159.7670	783.0113
k-means	179.8415	712.4325
k-means-CRO	160.2478	681.3785

TABLE II
SSE Value Obtained by FKM, FKM-CRO, K-Means and K-Means-CRO Clustering Algorithms in Three Real-Life Data Sets: Iris Wine and Wisc

SSE	iris	wine	wisc
FKM	55.4161	1848312.2104	11.6088e+13
FKM-CRO	53.8966	1780100.2589	8.5493e+13
k-means	71.7910	2441409.3106	21.7024e+13
k-means-CRO	60.2196	2212976.2357	15.2649e+13

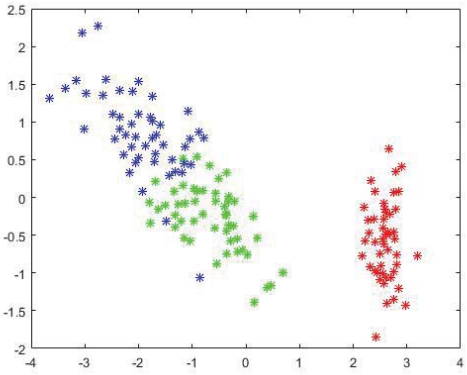
IV. Experiment

To validate the sensibility to initial value and efficiency of the presented algorithm, we conducted several experiments. The experimental results compare the EM-CRO algorithm with the k-means and FKM algorithm. The experimental datasets contain low latitude and high dimensional datasets. In order to facilitate the calculation of CRO, we used the Sammon technology to reduce the dimension.

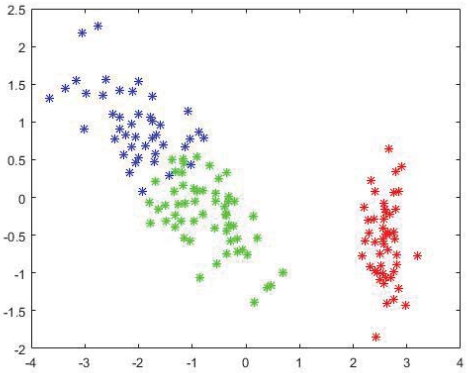
Provided datasets: Two artificial datasets: Data-60-2 and data-2000-2; three real-life data sets: Iris, wine and wisc. Data-60-2 is two-dimensional sparse data with 60 points and two non-overlapping clusters. Data-2000-2 is two-dimensional dense data with 2000 points and three clusters. Iris is a four-dimensional data with 150 points in three clusters. Wine is 178*14 datasets in three classes. Wisc is 699*10 data sets in two kinds.

For each experiment, we concluded the mean SSE value, derived that the result is stable. The algorithms are less sensitive to initial value than original algorithms.

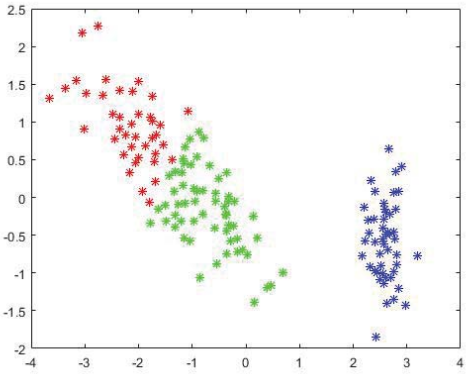
In the experimental process, Tables I and II express that CRO optimization algorithms of the k-means and FKM: k-means-CRO and FKM-CRO are better than the original algorithms, and the improvement of the k-means is greater than the FKM. This is because that k-means is more sensitive to the initialization than FKM. The comparing result of iris data in Fig. 1 shows that the clustering effect of EM-CRO is comparable to that of FCM, but it is better than k-means algorithm. Fig. 1, we used the Sammon technology to reduce



(a)



(b)



(c)

Fig. 1 Data (iris (a) data original partition(b) data partition using k-means-CRO and (c) data partition using fkm-CRO

dimensional processing. Despite some discrepancy, it can basically restore the original data attributes.

V. Conclusion

This paper provided a framework of chemical reaction algorithms of EM algorithm. This algorithm presented the combination between CRO operation and EM which makes algorithm efficient. In addition, structure of the algorithm can easily be controlled to find global minimum.

We compared the efficiency of the hybrid algorithm with EM algorithm on artificial datasets and real-life datasets. As shown in experiment section, we see that this algorithm is effective in solving optimization problems.

Recently, multiobjective clustering [3] has been an area of active research where a body of literature already exists. Our future work will focused on investigate multi-objective optimization on EM clustering and design to solve more problems well. Moreover high-dimensional data computation of the distances becomes a bottleneck, and the problem is the global algorithm's converging speed, we will study the fast global clustering algorithm in high-dimensional data, which can utilize the parallel [15] and artificial intelligence [3] methods, which are two main research directions.

Acknowledgment

This project is partially funded by the National Natural Science Foundation of China (Grant No. 61370095). The project is supported by Scientific Research Fund of Hunan Provincial Education Department (No. 13C333). The work was supported by research Project of the Education Department of Hunan Province (Grant No.14c0210).

References

[1] Albert Y. S. Lam and Victor O. K. Li, Chemical reaction optimization: a tutorial, Memetic Computing, Vol. 4, No. 1, 2013.
[2] Sanghamitra Bandyopadhyay, Genetic algorithms for clustering and fuzzy clustering, WIREs Data Mining and Knowledge Discovery, Volume 1, December 2011.
[3] Siripen Wikaisuksakul, A multi-objective genetic algorithm with fuzzy c-means for automatic data clustering, Applied Soft Computing, Vol. 24, 2014.
[4] David Hallac Jure Leskovec, Stephen Boyd, Network Lasso: Clustering and Optimization in Large Graphs, KDD15, August 2015.
[5] Grgoire A. Gallet and Fabio Pietrucci, Structural cluster analysis of chemical reactions in solution, The Journal of Chemical Physics, Vol 139, 2013.
[6] P. C. Srinivasa Rao1.Haider Banka1, Novel chemical reaction optimization based unequal clustering and routing algorithms for wireless sensor networks, Wireless Netw, Vol 12, January 2016.
[7] Yoon Mi Hamrick and Michael D. Morse, Comparative Cluster Reaction Studies of the V, Nb, and T a Series, The Journal of Physical Chemistry, Vol. 93, No. 17, 1989.
[8] M.-S. YANG, A Survey of Fuzzy Clustering, Mathl. Comput Modelling, Vol. 18, No. 11, 1993.
[9] Chenglong Tang Shigang Wang, Wei Xu, New fuzzy c-means clustering model based on the data weighted approach, Data&Knowledge Engineering, Vol. 69, 2010.
[10] Dan Li Hong Gu Liyong Zhang, A fuzzy c-means clustering algorithm based on nearest-neighbor intervals for incomplete data, Expert Systems with Applications, Vol 37, 2010.
[11] Matteo Brunelli Mario Fedrizzi Michele Fedrizzi, Fuzzy m-ary adjacency relations in social network analysis, Information Fusion, Vol 17, 2014.
[12] Du-Ming Tsai Chung-Chan Lin, Fuzzy C-means based clustering for linearly and nonlinearly separable data, Pattern Recognition, Vol 44, 2011.

[13] Lun Hu and Keith C. C. Chan, Fuzzy Clustering in a Complex Network Based on Content Relevance and Link Structures, IEEE Transactions on Fuzzy Systems, vol. 24, No 2, 2016.
[14] Burhan Tu rksen, Review of fuzzy system models with an emphasis on fuzzy functions, Transactions of the Institute of Measurement and Control , Vol 31, 2009.
[15] Isabel Timn, Jess Soto, Horacio Prez-Snchez, Jos M. Cecilia, Parallel implementation of fuzzy minimals clustering algorithm, Expert Systems With Applications, Vol 41, 2016
[16] Tien Trong Nguyen, ZhiYong Li, ShiWen Zhang, Tung Khac Truong, A hybrid algorithm based on particle swarm and chemical reaction, Expert Systems with Applications 41 (2014).
[17] Liang Bai, Jiye Liang, Cluster validity functions for categorical data: a solution-space perspective, Data Min Knowl Disc, 2014
[18] Raghu Krishnapuram,and James M. Keller, A Possibilistic Approach to Clustering, IEEE Transactions on Fuzzy Systems, vol. 1, no. 2, May 1993
[19] Dempster, A P, Laird, N M, Rubin, D B. Maximum Likelihood for Incomplete Data via the EM Algorithm (J). 1977.
[20] Hartigan J A, Wong M A. A K-means clustering algorithm.(J). Applied Statistics, 2013, 28(1):100-108.