International Journal of Smartcare Home
Vol.1, No.1 (2021), pp.45-60
https://dx.doi.org/10.21742/26531941.1.1.05

Research on Elastic Network Clustering Algorithm Based on Maximum Entropy

Liu Lu

Harbin Institute of Petroleum, Harbin, Heilongjiang 150027, China
77112428@qq.com

Abstract

In the information age, humans need to dig out potential information from a very large amount of information, and cluster analysis is one of the important methods. Most clustering analysis algorithms lack universal applicability, and they often encounter local minima problems when processing data sets with complex and diverse data structures. Cluster analysis has a strong ability to process uncertain information and is very robust. Among them, the elastic network algorithm, which belongs to unsupervised learning, has good geometric properties and can be solved for a specific objective function, which fits well with the definition of the clustering problem. Therefore, this paper has researched based on elastic network algorithm and applied the network in the field of cluster analysis. This paper proposes an Elastic Net of Clustering based on Maximum entropy (ENCM). According to the definition of clustering, the objective function of the elastic network algorithm is changed, and the principle of maximum entropy is used to determine the probability distribution of the data set without prior knowledge. Under the framework of the elastic network, the physical system is simulated, deterministic annealing technology is used to raise and lower the temperature of the system, control the network activity, and use the steepest descent method to track the minimum value.

Keywords: *Data analysis, Elastic network, Maximum entropy clustering algorithm*

1. Introduction

Data mining technology can effectively extract useful knowledge from a lot of messy information. This technology has been used in market analysis, business management, engineering design, scientific research, medical imaging, and other fields. Now that society has entered the information age, data is the main carrier of information. Making full use of and analyzing massive amounts of data, mining the hidden information and value in it, can provide more knowledge for decision-makers. Therefore, how to use specific techniques to analyze data and realize intelligent management of data has become a hot issue in the field of data mining. Many experts and scholars are also paying attention to such issues and have carried out many discussions and researches on them.

Cluster analysis can extract and analyze the characteristics of data from the data set, mine hidden laws, and achieve the purpose of classifying and summarizing data. It is a very important content in data mining technology. It has been used in research in multidisciplinary fields. After years of research, cluster analysis technology has developed very maturely and has also been applied in many industries. In biomedicine, the analysis of individual gene sequences is used

Article history:
Received (February 10, 2021), Review Result (March 15, 2020), Accepted (May 15, 2021)

eISSN: 2653-1941 IJSH

to gain an in-depth understanding of different groups, to provide doctors with important reference information to overcome certain diseases [1][2]. In terms of information retrieval, clustering algorithms can classify and summarize various information, facilitate subsequent maintenance, and help us obtain target information more quickly [3][4]. In terms of investment finance, through clustering analysis of various related information of different companies, the comprehensive situation of each company is objectively evaluated from the perspective of data, and more reliable suggestions and opinions are provided for investors [5][6]. In terms of meteorological prediction, clustering and analysis of a large amount of dynamic data provide important reference information for predicting climate change [7][8]. In terms of e-commerce, clustering is used to analyze customer background, behavior, and other information, to provide each customer with more personalized services [9][10], and to improve the service quality of the platform. In addition, cluster analysis also has many applications in statistics, artificial intelligence, and other fields, such as decision analysis, computer vision, and so on.

Elastic network algorithm is a kind of neural network, which is mainly used in combinatorial optimization, information processing, and other fields, and belongs to unsupervised learning. The operation of the network does not require any manual training. In addition, the elastic network algorithm also has the advantages of strong robustness and stable operation. Its geometric properties can fit well with the definition of clustering. In the process of clustering, it can jump out of the local minima to some extent and find a better solution. This paper combines the theory and characteristics of cluster analysis and elastic network algorithm, applies the elastic network solving model to the clustering problem, and proposes a clustering algorithm based on elastic network. In addition, given the common problems of noise interference and poor results of high-dimensional data in cluster analysis, an elastic network clustering algorithm with weighted characteristics is proposed to fill up the defects of traditional clustering algorithms.

The more mainstream clustering algorithms include hierarchical clustering algorithms, partition-based clustering methods, grid-based clustering algorithms, density-based clustering algorithms, and model-based methods. Clustering technology is not only the above five types, there are currently many new clustering algorithms, such as granularity-based, entropy-based, uncertain clustering, spectral clustering, kernel clustering, and other methods [11][12]. Although there are many types of clustering algorithms, almost every algorithm is designed for one or several problems, so these algorithms still have some problems during operation [13]. Such as the processing of interference values such as noise in the data, and the reasonable explanation of the final clustering scheme. Through investigation, there are many types of research and directions for cluster analysis technology. For example: determining appropriate similarity measures, preprocessing of data sets, interpretability of clustering results, sensitivity to the order of input data, etc.

2. Neural networks

Artificial Neural Networks (ANN), also known as Neural Networks (NNs) or Connection Model. It is defined as a structure composed of densely interconnected elements (called Artificial neurons or nodes) that can be adapted to simply process data. These elements achieve the purpose of data processing and knowledge representation through large-scale parallel computing of data. The artificial neural network has the characteristics of non-linearity, high parallelism, robustness, learning ability, processing inaccurate and fuzzy information, and generalization ability. Similar to the biological nervous system, the most basic building block

of ANN is the artificial neuron. Different neural networks can be obtained by connecting multiple neurons in different ways.

The Error Back Propagation (BP) algorithm is a feed-forward network, which contains multiple levels and was originally proposed by Werbos. However, because the research on neural networks was in a low period at that time, the algorithm was not paid attention to. About ten years after the algorithm was proposed, a team of scientists led by Rumelhart and McCelland re-proposed the theory and developed it. The basic idea of the BP algorithm is: after the samples that need to be learned are input into the network, the weight and deviation of the training network are adjusted through the method of backpropagation so that the final result is as close to the present value as possible. There are two training processes for the BP network, namely forward propagation of information and backpropagation of error. Its model structure includes an input layer, hidden layer, and output layer, as shown in [Figure 1]. There are two situations for ending network training: one is that the difference between the value of the network output and the expected value is reduced to the allowable range; the other is that the number of network learning reaches the previously set value.

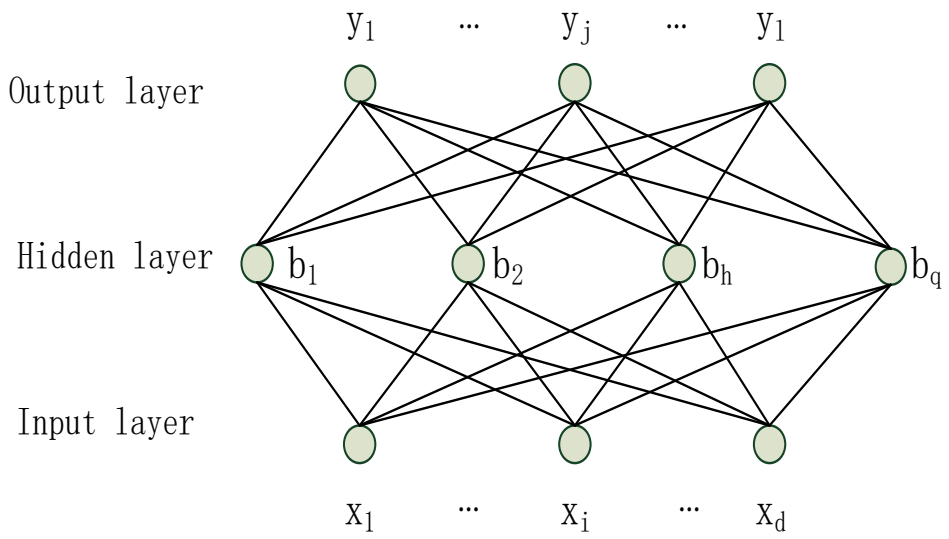

Figure 1. BP network structure

The BP algorithm is currently the most successful ANN algorithm. It has strong nonlinear mapping capabilities and can easily perform large-scale parallel data processing. It can also store information in a distributed manner and has good self-learning capabilities. Although the middle layer of the BP network can be replaced flexibly, there is currently no theory to guide the selection of the middle layer and the number of units. Moreover, the algorithm requires prior knowledge, and the training time is too long. Problems such as complete inability to train due to parameter discomfort, unstable network learning and memory, and easy to fall into local minimums.

The scholar GA Carpenter of Boston University in the United States has been committed to establishing a unified mathematical theory of human psychology and cognitive activities. Adaptive Resonance Theory (ART) is the core part of this theory, and subsequently proposed the ART network with S. Grossberg. ART network is an important representative of competitive learning, which is composed of a comparison layer, a recognition layer, a

recognition threshold, and a reset module. The original ART network can only process Boolean input data, often called ART1, and its structure is shown in [Figure 2]. After research and development, an extended form of ART1 appeared-it can process continuous analog signals ART2, a hierarchical search model ART3 compatible with ART1 and ART2 structures, and a combination of fuzzy ART and other algorithms for fuzzy processing. The ART network alleviates the "Stability-plasticity dilemma" and can adapt to new input modes. Learn the new knowledge while keeping the memory of the old knowledge to avoid the modification of the old model. Therefore, you can do incremental learning or online learning, which is also the biggest advantage of the ART network.

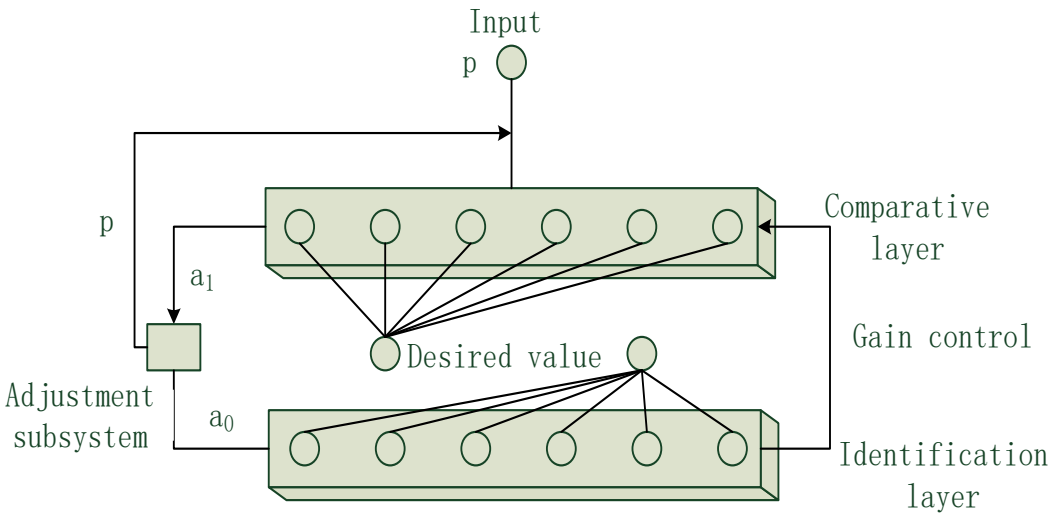

Figure 2. ART-1 network structure

Kohonen believes that for different signals in a neural network, different neurons will respond to them in the most appropriate way to form a feature map or ordered map with topological significance. Therefore, Kohonen proposed the Self-Organizing Map (SOM) network in 1982. This model can map input data with higher dimensions to space with lower dimensions (usually two-dimensional) while maintaining the topological structure of the data. It is a competitive learning network model. The biggest advantage of SOM is that it has self-learning ability, that is, without prior knowledge, it can adaptively change the network structure and parameters, and achieve the purpose of automatically finding the inherent laws in the sample.

Since the SOM network does not require prior knowledge, it can easily find the optimal solution through self-learning to find adaptive weights, and it has the characteristics of maintaining topological structure, maintaining probability distribution, and visualization. It has attracted the attention of many scholars in recent years. Has also been widely used in image processing, classification, clustering, data analysis, combination optimization, prediction, and other fields. However, SOM also has disadvantages, because the data, in reality, does not always have convex characteristics. When the initial conditions of the network are poor, the network tends to fall into a local minimum.

The elastic network is an extension of the SOM network. In addition to unsupervised learning, maintaining topological structure, visualization, and other characteristics, it also has good geometric properties. It can automatically solve the objective function, without manual

guidance and training, and ensure that the network can converge globally. The clustering problem can be boiled down to a problem of solving minimum values. Therefore, this paper uses the elastic network as the basic model to study the problems in the field of cluster analysis and explore a new algorithm that can improve the quality of clustering.

3. Cluster analysis algorithm

3.1. Overview of clustering

Clustering analysis distributes the sample objects in the data set according to certain rules so that in the assigned classes, the elements of the same class are as similar as possible, and the elements between different classes are as different as possible. The general mathematical model of clustering can be described as: Given the sample set $X = \{x_1, x_2 \dots x_i \dots x_n\}$ to be studied, x_i represents the i-th sample in the sample set, and each sample has m feature attributes, $X_i = \{x_{i,1}, x_{i,2} \dots x_{i,l} \dots x_{i,m}\}$. The goal is to find k clusters $C = \{C_1, C_2 \dots C_i \dots C_k\}, (k < n)$ in the sample set, then these k clusters satisfy the condition:

$$C_i \neq \emptyset, i = 1,2, \dots, k \ ;$$

$$C_1 \cup C_2 \cup \dots \cup C_k = X \ ;$$

$$C_i \cap C_j = \emptyset, i, j = 1,2, \dots, k \ and \ i \neq j$$

According to the above conditions, each element in the sample set will be assigned to one class, and can only be assigned to one class. To cluster the sample set, the traditional clustering model can be boiled down to solving the minimization problem of Equation 1 [14]. Among them, $y_1, y_2 \dots y_k$ are the cluster centers. This problem is a non-convex optimization problem. Many scholars have studied the method of finding the global optimal solution to this problem, but they have not yet found a very suitable solution.

$$minD = \sum_{j} \sum_{x_i \in C_j} \left(d(x_i, y_i)\right)^2 \quad (1)$$

3.2. Main methods of clustering and comparison

Cluster analysis technology can divide samples and form clusters based on the relationship between the actual significance and mathematical characteristics of the sample objects. The purpose is to make the samples in the same cluster as similar as possible, or even identical. The similarity between samples in different clusters is as low as possible, or even completely different [15]. The essence of the clustering algorithm is to enable the computer to automatically group samples according to certain criteria, and there is a certain approximate relationship between the samples in each group. Among the clustering analysis algorithms that have been developed today, some algorithms need to be learned and trained before they can be used, such as semi-supervised and supervised clustering algorithms. However, most clustering algorithms do not need training, that is, no labeled data sample set is needed to train the network. Through the joint efforts of experts and scholars, the proposed clustering algorithm can be divided into five categories: hierarchy, division, density, grid, and model. This section will elaborate on the concepts, principles, advantages, and disadvantages of various algorithms.

The basic idea of the hierarchical clustering algorithm is to allocate data samples from the hierarchical aspect to form a tree-like clustering structure. Generally, such algorithms are more

flexible and require fewer parameters or even no need for parameter adjustment. This type of algorithm can be divided into two types according to the process of cluster formation, one is an aggregation (bottom-up), and the other is split (top-down). In the process of each iteration, the layer-based method will use a greedy strategy to find the most suitable solution to condense or split. Although this method can find clusters of various shapes, the computational complexity is relatively high and the running cost is relatively high. In addition, this method is sensitive to noise values, and there is no uniform standard for the conditions under which the algorithm stops running. Typical hierarchical clustering algorithms include BIRCH, CHAAMELEON, CURE, CLUBS, and so on.

Clustering algorithm based on partition usually uses distance as the standard for clustering [16], and iteratively converges on a given data set according to the principle of minimum distance, to achieve partition and obtain target clusters without crossover. A typical and common example of such methods is k-means. The k-means algorithm is simple and fast and has been applied in many fields. K-means has a very good clustering effect on data sets with convex characteristics. However, k-means randomly selects points to initialize the cluster centers. This method of initializing cluster centers makes k-means very sensitive to outliers and outliers. The results of multiple clustering for the same problem are not unique. Later, someone proposed the k-medoids algorithm, which changed the way of initializing cluster centers to randomly select k sample points as the initial cluster centers. This move weakened the impact of noise data on clustering to a certain extent, but did not solve the problem of not unique clustering results for the same problem, and also increased the computational cost of the algorithm. In addition, real data sets do not always have convex characteristics, so methods such as k-means and k-medoids often fall into local minima during operation. Many experts and scholars have made various improvements and optimizations to k-means, such as k-means. In addition to a series of improved algorithms based on k-means, partition-based clustering methods also include CLARA, FCM, and so on.

The grid-based clustering method is to first form a grid structure in the data space, which contains several finite cells, and then cluster the samples in units of cells. This type of algorithm maps the samples to cells one by one according to a certain mapping function. Determine the density of this cell by counting the number of cell samples, and finally group the densely connected cells into one category. The processing speed of the grid-based method is generally related to how many cells the sample space is divided into, and has nothing to do with the number of samples to be processed. Therefore, this type of method relies heavily on parameters, and the quality of the clustering results and the efficiency of the algorithm are constrained by each other, and they need to be weighed when used. In addition, for data sets with uneven distribution of data samples, the results of clustering using this method are not ideal, which is also its limitation. Typical grid-based clustering algorithms include CLIQUE, STING, Wave Cluster, etc.

The density-based clustering algorithm is an algorithm that uses distance as a measurement standard and can only find clusters with convex characteristics, that is, "quasi-circular" clusters. Some people have proposed clustering based on the density of data samples. This type of algorithm judges the continuity between samples by sample density, and expands and clusters according to the connectable samples, and divides the data space according to the density, that is, the density is high. The area of ?? is a category, and the area of low density is a category, and the solution is finally obtained. This method can find clusters of various shapes in the data sample, but therefore, determining the "sample neighborhood" and quantifying high-density areas have become the key issues of this type of algorithm. DBSCAN is a well-known density-based algorithm. This method describes the closeness of samples according to a set of

"neighborhood" parameters. The density reachability relationship can derive the density-connected sample set, and define the largest among them. As a cluster [17].

Model-based clustering algorithms usually try to classify each cluster. Suppose a model is determined, and then find a data set that fits the model. Such algorithms usually assume that the data set to be processed follows a certain probability distribution. There are two main attempts: the statistical scheme and the neural network scheme. A neural network is a mathematical model, and its basic component is a simple processing unit (or neuron). Connect these neurons to realize the structure and function similar to the human brain, which can realize functions such as large-scale parallel processing and computing and distributed storage. The neural network is very robust and can remove interference factors such as noise. Therefore, using this model in the field of clustering may be expected to solve the problem of noise in traditional clustering problems. At present, there are two main methods for introducing neural networks into the clustering field: ART (Adaptive Resonance Theory) and SOM (Self-Organization Map) [18].

Traditional clustering analysis algorithms are generally divided into five categories. Since these algorithms strictly divide the data to be processed according to certain rules, they can be collectively referred to as hard clustering or hard partitioning. With the emergence of fuzzy technology, algorithms for applying this technology to the field of clustering have also been proposed. This method can be called fuzzy clustering. Unlike hard clustering, fuzzy clustering does not divide the data samples into a certain class but forms an uncertain relationship between the sample and the class. You can make a sample belong to several classes at the same time, and then describe the relationship between the sample and its subordinate classes. Compared with hard clustering, the processing method of fuzzy clustering on data information is more compatible with the data model in the real world.

Many traditional clustering algorithms are based on distance, but there is a lot of irrelevant information in the space with higher dimensions. The distribution of the data is also sparser than in the low-dimensional space, and the hidden clusters are also smaller. This makes the original method unable to obtain a better clustering scheme in a high-dimensional space. Currently, the methods used to solve clustering in high-dimensional spaces usually include molecular spatial clustering methods, bi-clustering methods, and dimensional reduction methods.

4. Elastic network clustering algorithms based on maximum entropy

4.1. Related technical principles

The concept of entropy was first proposed by the German physicist Clausius in 1865, and then Shannon proposed information entropy in 1948 and applied it as a measurement tool in thermodynamics. The proposal of these concepts has inspired many scientists to extend the concept of entropy to many other fields, resulting in the concept of thermodynamic entropy and generalized entropy [19]. The generalized entropy covers a wide range of meanings, including fuzzy entropy, topological entropy, conditional entropy, meteorological entropy, economic entropy, and so on. It is used to describe the uncertainty, degree of mixing, and disorder of things, or a measure to describe the state of certain material systems or the extent to which certain material system states may appear.

To describe the entropy H from the perspective of probability theory, it can be assumed that there are n mutually independent selection results in the samples $a_1, a_2, \ldots, a_n$, and the

probability of each of them is $p, p_2, \ldots, p_n$, then the degree of uncertainty H of the event can be described by Equation 2.

$$H = -C\sum_{i=1}^{n} p_i \, ln(p_i) \tag{2}$$

When $p_1 = p_2 = \cdots = p_n$, $p_i = 1/n$, then

$$H = -C ln^n$$

It is worth mentioning that the principle of maximum entropy is universal and can be used in both classical mechanics and quantum mechanics. Both equilibrium and non-equilibrium states can be handled. In addition, this principle satisfies the first principle and consistency requirements, that is, the principle of maximum entropy uses the least information to obtain the most detached (Maximally noncommittal) solution, and the solution obtained is independent of the solution step.

Deterministic annealing technology is an important part of Physical Computation [20], which was originally proposed by K. Rose, a Ph.D. from the California Institute of Technology in 1990 [21]. The so-called annealing is a natural process, and the deterministic annealing technology is proposed based on this natural process. Before introducing the annealing process, first, introduce the melting process. When heating a solid substance, the particles in the substance are originally in an orderly crystalline state. As the temperature rises, it deviates more and more from its equilibrium position. When the temperature rises to the dissolution temperature, the original movement rules of the particles in the solid will be destroyed, and the substance will turn from the solid to the liquid. There is no order in the arrangement of the particles in the liquid state. The particles in the matter can move freely and are evenly distributed in all directions. The purpose of the melting process is to remove the uneven state that may exist in the system.

The deterministic annealing technique is to assumes the system as a physical model. The temperature changes slowly from high to low. Under the state of each temperature, the minimum value of the free energy function is searched through a certain optimization method, and the global optimal point of the free energy function is tracked until the temperature drops to zero. The simulated annealing algorithm uses the Metropolis sampling criterion to simulate the equilibrium state of the system. Therefore, the solution speed of the simulated annealing algorithm is not as fast as the deterministic annealing technology for the same problem.

The steepest descent method proposed by the mathematician Cauchy in 1847, also known as the gradient method, is a method for solving unconstrained optimization problems. This method uses the negative gradient direction as the search direction and can search quickly and finely locally. More importantly, since the direction of the negative gradient is the local steepest descent direction of the function, this method can still converge to the local optimal value even from a bad initial point.

In the process of running iterations of this method, first, select an appropriate α, α represents the step length of each iteration of the program. It can be seen that $k\alpha$ is the minimum value of Formula 3, as shown in Formula 4. In each iteration, starting from the $k(x)$ point, move a certain distance along the opposite direction of the gradient to gradually reduce the objective function, and finally, find the minimum point.

$$\emptyset_k(\alpha) = f\left(x^{(k)} - \alpha\nabla f\left(x^{(k)}\right)\right) \tag{3}$$

$$\alpha_k = argminf\left(x^{(k)} - \alpha\nabla f\left(x^{(k)}\right)\right), \alpha \geq 0 \quad (4)$$

The most obvious difference between the steepest descent method and the gradient descent method is that the direction of change of $x^{(k)}$ is selected. As in Formula 5, Δ_{nsd} represents the direction of the fastest descent. If $\|v\|$ is the Euclidean norm, then the two descent methods are the same.

$$\Delta_{nsd} = argmin_v(\nabla f(x)^T v \quad \|v\| \leq 1) \quad (5)$$

But in fact, the gradient is not fixed, it is always changing, and each gradient descent is assumed to be fixed during the iteration process, that is, the gradient direction is only the gradient direction of the starting point $x^{(k)}$. The gradient direction does not represent the gradient direction of other points between the initial point and the endpoint, that is, the gradient directions of other points between the initial point and the endpoint may be different. The Euclidean norm is only one of many calculation methods, and it is not suitable for all situations. Only by choosing a proper norm for a specific problem can we get a faster descent direction.

The steepest descent method is the simplest among many unconstrained optimization methods. This method has relatively low requirements for the initial point, can guarantee the overall convergence of the system, and requires less calculation and storage space. Big. The cluster analysis is a non-convex optimization problem, so this paper chooses the steepest descent method to gradually optimize the objective function.

4.2. The original elastic networks

In 1987, Durbin and Willshaw proposed a heuristic combinatorial optimization algorithm—The elastic net method. Used to solve the TSP problem (Traveling salesman problem). This algorithm has good geometric characteristics and is very suitable for solving problems with the optimal solution of a unique objective function. It can solve the objective function automatically without manual guidance. This section will explain the principles of ENA in detail.

The elastic network algorithm will describe each city point in multivariate space, and generate a closed circle with elastic points centered on the center of the data set in the space formed by the city points. This ring is called a Rubber band. As the network continues to iterate, the elastic band is deformed, and the position of the elastic node is constantly changing, and it is constantly approaching the city point. Until all city points are covered by flexible nodes, the network reaches a stable state, and it is considered that an approximate solution to the TSP problem is obtained. When running iterations, the elastic network uses two forces to pull changes in elastic nodes. One kind of force is the attraction of the city point adjacent to the elastic node i to the elastic node i, making the elastic node i gradually approach the city point. One kind of force is the tension of the elastic node adjacent to the elastic node i to the elastic node I, forcing the distance between the adjacent elastic nodes to be the shortest.

Assuming that a given data set $X = \{x_1, x_2 \ldots x_i \ldots x_n\}$ contains n city points, the elastic band $Y = \{y_1, y_2 \ldots y_m\}$ with m elastic nodes is initialized in the space, generally, $n < m$. As the network iterates, the elastic network tracks the minimum value of the energy function formula 6:

$$E = -\alpha K \sum_{i=1}^{n} ln \sum_{j=1}^{m} \emptyset(|x_i - y_i|, K) + \beta \sum_{j=1}^{m} \left|y_{j+1} - y_j\right|^2 \quad (6)$$

$$\emptyset(d,K) = e^{-\frac{d^2}{2K^2}}$$

Among them, K gradually decreases, which corresponds to the temperature in the deterministic annealing method. The parameters α 、 β respectively determine the degree of influence of the city point on the force of the elastic node and the force between the elastic nodes. The amount of change required by the elastic node in each iteration is Formula 7:

$$\Delta y_j = -K\frac{\partial E}{\partial y_j}$$

$$= \alpha\sum_{i=1}^{n}\omega_{ij}(x_i - y_j) + \beta K(y_{j+1} - 2y_j + y_{j-1}) \tag{7}$$

Among them, ω_{ij} represents the degree of influence of the city point on the elastic node j on the path, and its description is shown in formula 8.

$$\omega_{ij} = \frac{e^{-\frac{|x_i - y_i|^2}{2K^2}}}{\sum_{j=1}^{m} e^{-\frac{|x_i - y_i|^2}{2K^2}}} \tag{8}$$

When $K \to 0$, the energy function obtains a local minimum, and each city point matches at least one elastic node. When the city point has and only matches one flexible node, the network is considered to obtain the optimal solution.

4.3. Elastic network clustering algorithm based on maximum entropy

For the data set $X = \{x_1, x_2 \ldots x_i \ldots x_n\}$ with n data points in the m-dimensional space, the target is divided into k clusters. Initialize an elastic band $Y = \{y_1, y_2 \ldots y_k\}$ with k elastic nodes in the same space. Define the energy function of the elastic network as:

$$E = \sum_j \sum_j p_{i,j} |x_i - y_i|^2 \tag{9}$$

Among them, $p_{i,j}$is the probability distribution of the system. Since there is no previous one for reference, the specific probability distribution pattern of the system cannot be determined at will. If the clustering problem is regarded as a model of a physical system, the data points are analogous to molecules in the physical system, and the entropy function H is defined for the fixed elastic node y_i as:

$$H = -\sum_{i=1}^{n}\sum_{i=1}^{n} p_{i,j}\, ln\, p_{i,j} \tag{10}$$

Using the principle of maximum entropy, under the constraint formula 9, by solving the variational problem, the probability distribution of the data point x_i belonging to a certain cluster is:

$$p_{i,j} = \frac{e^{-\beta\times\frac{|x_i - y_i|^2}{2}}}{\sum_j e^{-\beta\times\frac{|x_i - y_i|^2}{2}}} \tag{11}$$

In addition, according to the characteristics of the elastic network itself, the consideration of the influence between the elastic nodes can be described by Equation 12.

$$E_{nod} = \frac{1}{2}\lambda\left|y_j - y_{j-1}\right|^2 \tag{12}$$

Among them, the constant λ is the elastic coefficient, and the free energy function corresponding to the energy function is:

$$F = -\frac{1}{\beta}\sum_i ln\sum_j e^{-\beta\times\frac{|x_i-y_i|^2}{2}} + \frac{1}{2}\lambda\left|y_j - y_{j-1}\right|^2 \tag{13}$$

Then the amount of change required for each iteration of the elastic node is:

$$\Delta y_j = -\Delta\tau\frac{\partial F}{\partial y_j}$$

$$= \Delta\tau\sum_i^n P_{ij}\left(x_i - y_j\right) + \Delta\tau\left(y_{j+1} - 2y_j + y_{j-1}\right)\forall j \tag{14}$$

Among them, the constant $\Delta\tau$ represents the time step, and an appropriate value needs to be selected to make the network run fully.

This chapter combines the principle of maximum entropy and the idea of simulated annealing constructs a new energy function based on the elastic network model, and uses the principle of maximum entropy to determine the probability distribution of data points, making the algorithm more universal. The flowchart of the ENCM algorithm is shown as in [Figure 3].

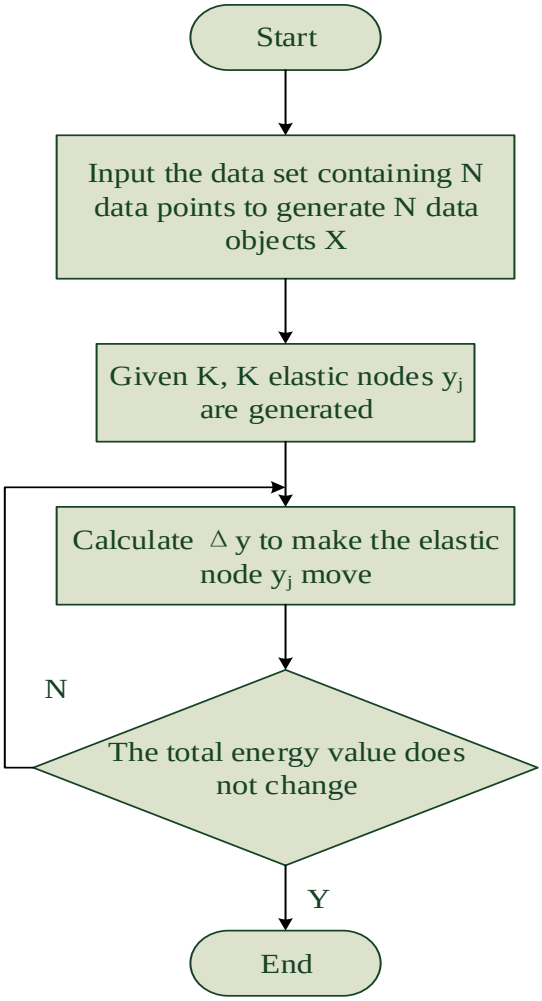

Figure 3. The algorithm flow chart

4.4. Experiment and simulation of ENCM algorithm

To verify whether this method can effectively perform clustering, this paper uses the evaluation standard SED value to evaluate and test data sets of different dimensions and different magnitudes. The calculation formula of the SED value is shown in Equation 15.

$$SED = \sum_{i=1}^{n} \sum_{j=1, x_i \in c_j}^{k} d(x_i, y_j) \tag{15}$$

The algorithm parameters of this experiment are set as $\beta = 3$, $\Delta\tau = 0.001$, $\lambda = 0.5$, the center point of the elastic band at the beginning is the center point of the data set, and the radius $R = 0.45$. To verify whether ENCM improves the quality of clustering results. This paper uses the classic k-means, k-medoids, and Birch algorithms to conduct comparative experiments. Since k-means and k-medoids are affected by the position of the initial center point, the results of multiple runs of the program on the same problem are not fixed, so use each data set separately.

Table 1. Comparison of clustering results of random data sets

Dimension	Number	Cluster	SED			
			k-means	k-medoids	Birch	ENCM
2D	100	3	45.34	46.22	43.32	40.98
		5	49.12	42.43	36.22	34.86
	500	3	421.67	429.89	325.56	322.45
		5	201.08	226.12	278.12	245.19
	1000	3	717.90	712.87	508.67	432.99
		5	561,98	583.45	413.37	389.56
3D	100	3	79,76	83,91	76.23	74.09
		5	68,24	70.87	62.56	60,77
	500	3	622.78	634.12	531.08	505.88
		5	568,25	573.66	588.34	499.21
	1000	3	1066.45	1233.56	917.04	913,65
		5	903.44	950.87	830.11	810.14

Table 2. Comparison of clustering results of high-dimensional random data sets

Dimension	Number	Cluster	SED			
			k-means	k-medoids	Birch	ENCM
7D	100	3	89.14	91.02	83.32	80.83
		5	87.92	80,43	76.29	74.06
	500	3	621.67	629.89	525.56	522.45
		5	501.08	526.12	478.12	445.19
	1000	3	1488,98	1576.09	1308.11	1300.78
		5	1261,98	1383.45	1213.37	1189.56
10D	100	3	99,76	103,91	96.23	94.09
		5	88,24	90.87	82.56	80,77
	500	3	722.78	734.12	631.08	605.88
		5	768,25	773.66	688.34	599.21
	1000	3	1766.45	1833.56	1617.04	1513,65
		5	1603.44	1650.87	1530.11	1410.14

The two methods were tested 20 times, and the average value of SED was taken as the final result for comparison. Table 1 lists the results of the above four methods and compares them. The ENCM algorithm can obtain better clustering results than k-means, k-medoids, and Birch.

To study the clustering effect of the ENCM algorithm in a higher-dimensional space, this paper selects 7-dimensional and 10-dimensional data for experiments and selects 100, 500, and 1000 random data sets with uneven distribution in each dimension. The three methods of k-means, k-medoids, and Birch are still used for comparison, and the experimental results are shown in Table 2. It can be seen from the table that for clustering in random data sets with higher dimensions, the clustering scheme obtained using the ENCM algorithm is still better than k-means, k-medoids, and Birch.

5. Conclusion

This article first summarizes the research status of cluster analysis, neural network, and elastic network at home and abroad. Then the article introduces the definition of clustering and several classic methods in cluster analysis. Traditional clustering algorithms can be divided into five types. Different algorithms are designed for different problems. Each clustering algorithm

has its advantages and disadvantages. Aiming at the problems of clustering results in the clustering algorithm, such as susceptibility to noise interference and poor quality of multi-dimensional spatial clustering. This paper studies the clustering problem based on the elastic network model and proposes an elastic network clustering algorithm ENCM and a weighted elastic network clustering algorithm WENC. Compared with traditional clustering algorithms, this paper chooses an elastic network model suitable for solving the objective function. Because the original elastic network cannot be directly used to solve the clustering problem. Based on the definition of the clustering problem and the characteristics of the elastic network model, this paper combines the principle of maximum entropy, simulated annealing, and the steepest descent method. Modify the objective function of the elastic network to solve the clustering problem, and verify its clustering performance through a random data set. It turns out that the effect of the elastic network clustering algorithm is better than traditional k-means and k-medoids.

References

[1] Y. Hui, "Cluster analysis of gene expression data," Nanjing University of Aeronautics and Astronautics, **(2007)**

[2] S. Zhao, H. Radner, D. Thong, S. J. Duffield, D. M. Hughes, and R. J. Moots, "Multimorbidity in patients with axial spondyloarthritis: A cluster analysis," Rheumatology, vol.58, no.4, pp.37-38, **(2019)**

[3] R. V. Pamba, E. Sherly, and K. Mohan, "Self-adaptive frequent pattern growth-based dynamic fuzzy particle swarm optimization for web document clustering," In Computational Intelligence: Theories, Applications, and Future Directions-Volume II: Springer, vol.799, pp.15-25, **(2019)**

[4] G. Kou and C. Lou, "Multiple factor hierarchical clustering algorithm for large scale web page and search engine clickstream data," Annals of Operations Research, vol.197, no.1, pp.123-134, **(2012)**

[5] X. Guibiao and X. Jiaoli, "The application of factor analysis and cluster analysis in financial industry stock investment," Coastal Enterprises and Technology, no.4, pp.11-16, **(2016)**

[6] H. Li, S. Zhao, Q. Sun, and Q. Li, "Construction and analysis of financial security index evaluation system based on weighted principal component distance clustering," Mathematics in Practice and Knowledge, no.1, pp.90-102, **(2018)**

[7] J. A. Acero, E. J. Koh, G. Pignatta, and L. K. Norford, "Clustering weather types for urban outdoor thermal comfort evaluation in a tropical area," Theoretical and Applied Climatology, vol.139, no.1-2, pp.659-675, **(2020)**

[8] P. H. Thong, "Some novel hybrid forecast methods based on picture fuzzy clustering for weather nowcasting from satellite image sequences," Applied Intelligence, vol.46, no.1, pp.1-15, **(2017)**

[9] G. P. Papamichail and D. P. Papamichail, "The k-means range algorithm for personalized data clustering in e-commerce," European Journal of Operational Research, vol.177, no.3, pp.1400-1408, **(2007)**

[10] U. Rahardja, T. Hariguna, and W. M. Baihaqi, "Opinion mining on e-commerce data using sentiment analysis and k-medoid clustering," In 2019 Twelfth international Conference on Ubi-Media Computing (Ubi-Media): IEEE, pp.168-170, **(2019)**

[11] A. Saxena, M. Prasad, A. Gupta, N. Bharill, P. P. Patel, and A. Tiwari, "A review of clustering techniques and developments," Neurocomputing, vol.267, pp.664-681, **(2017)**

[12] T. Zhou and H. Lu, "Research progress of clustering algorithms in data mining," Computer Engineering and Applications, vol.48, no.12, pp.100-111, **(2012)**

[13] S. Solorio-Fernández, J. A. Carrasco-Ochoa, and J. F. Martínez-Trinidad, "A review of unsupervised feature selection methods," Artificial Intelligence Review, vol.53, no.2, pp.907-948, **(2019)**

[14] K. Rose, E. Gurewitz, and G. C. Fox, "Statistical mechanics and phase transitions in clustering," Physical Review Letters, vol.65, no.8, pp.945-948, **(1990)**

[15] Z. Sun, "Research and application of high-dimensional data clustering algorithm," Jiangnan University, **(2017)**

[16] K. Krishna and M. Narasimha Murty, “Genetic k-means algorithm,” IEEE Transactions on Systems Man and Cybernetics-Part B Cybernetics A Publication of the IEEE Systems Man and Cybernetics Society, vol.29, no.3, pp.433-439, **(1999)**

[17] Z. Zhou, “Machine Learning,” Airport, no.2, pp.94-94, **(2018)**

[18] T. Kohonen, “Self-organized formation of topologically correct feature maps,” Biological Cybernetics, vol.43, no.1, pp.59-69, **(1982)**

[19] G. Han, “Entropy and the application of entropy,” Journal of Inner Mongolia Normal University (Education Science Edition), no.4, pp.13-15, **(2001)**

[20] G. C. Fox, “Physical computation,” Concurrency: Practice and Experience, vol.3, no.6, pp.627-653, **(1991)**

[21] K. Rose, E. Gurewitz, and G. C. Fox, “Statistical mechanics and phase transitions in clustering,” Physical Review Letters, vol.65, no.8, pp.945-948, **(1990)**

This page is empty by intention.