International Journal of Multimedia and Ubiquitous Engineering
Vol.10, No.12 (2015), pp.255-264
http://dx.doi.org/10.14257/ijmue.2015.10.12.26

The Research and Application of Reservoir Identification Model Based on Smap-ED

Huoyou Li[1], Jia Zhao[2], Dongyun Lin[2], Jingsong Su[3],
Qingqiang Wu[3] and Qingqi Hong[3*]

[1]*School of Information Engineering, Longyan University*
[2]*Information Science and Engineering School, Xiamen University*
[3]*Software School, Xiamen University*
**hongqq@xmu.edu.cn*

Abstract

In order to provide a scientific decisions for prospect new oil formation and reduce production costs, this paper proposed reservoir identification model based on Smap-ED. Using the actual seismic attribute data as the research object, the principles are analyzed to explore the reservoir recognition model , Through the experimental analysis of model of effectiveness evaluation indexes , the five models of applicability and effectiveness was comparative study from the relative index, external index and running time. The experimental results show model based on Smap-ED of seismic attribute data clustering effect is better than model based on S-Map and the running time of algorithm close to model based on SOM, this model can provide more effective support for scientific decision-making.

Keywords: *reservoir identification; Smap-ED; evaluation indexes*

1. Introduction

Oil deposit exploration industry is a industry that have enormous data information[1], the research on the dispose of seismic attribute data is closely related to the requirement and the evolution of oil exploration and development, so we need to use the subtle reservoir description, discriminatory treatment, classified implementation, and we can unearth the potential of oil well specifically, exploit each oil well scientifically and reasonably, then can achieve the optimization of social and economic benefit [2]. Cluster analysis is a effective data mining method that can achieve distinguishing the geological categories.

SOM form topological mapping from the input layer to output layer, now it has achieved a great success in many fields [3]. However, this algorithm is still essentially existing the following deficiencies [4]: A density model is not defined in data space; the self-organizing feature map training process is not working by optimizing the objective function, learning training process cannot ensure convergence.

GTM model clustering effect for seismic attribute data is better than SOM, but the running time of this algorithm costs longer than SOM. GTM-ED model clustering effect is slightly better than the original GTM model, and the running time of algorithm is within the acceptable range

For seismic attribute data, S-map can get a better clustering effect, and the running time of the algorithm is close to SOM, so it has realized the high efficiency and high quality of clustering.

[*] Corresponding Author

ISSN: 1975-0080 IJMUE

2. Methodology

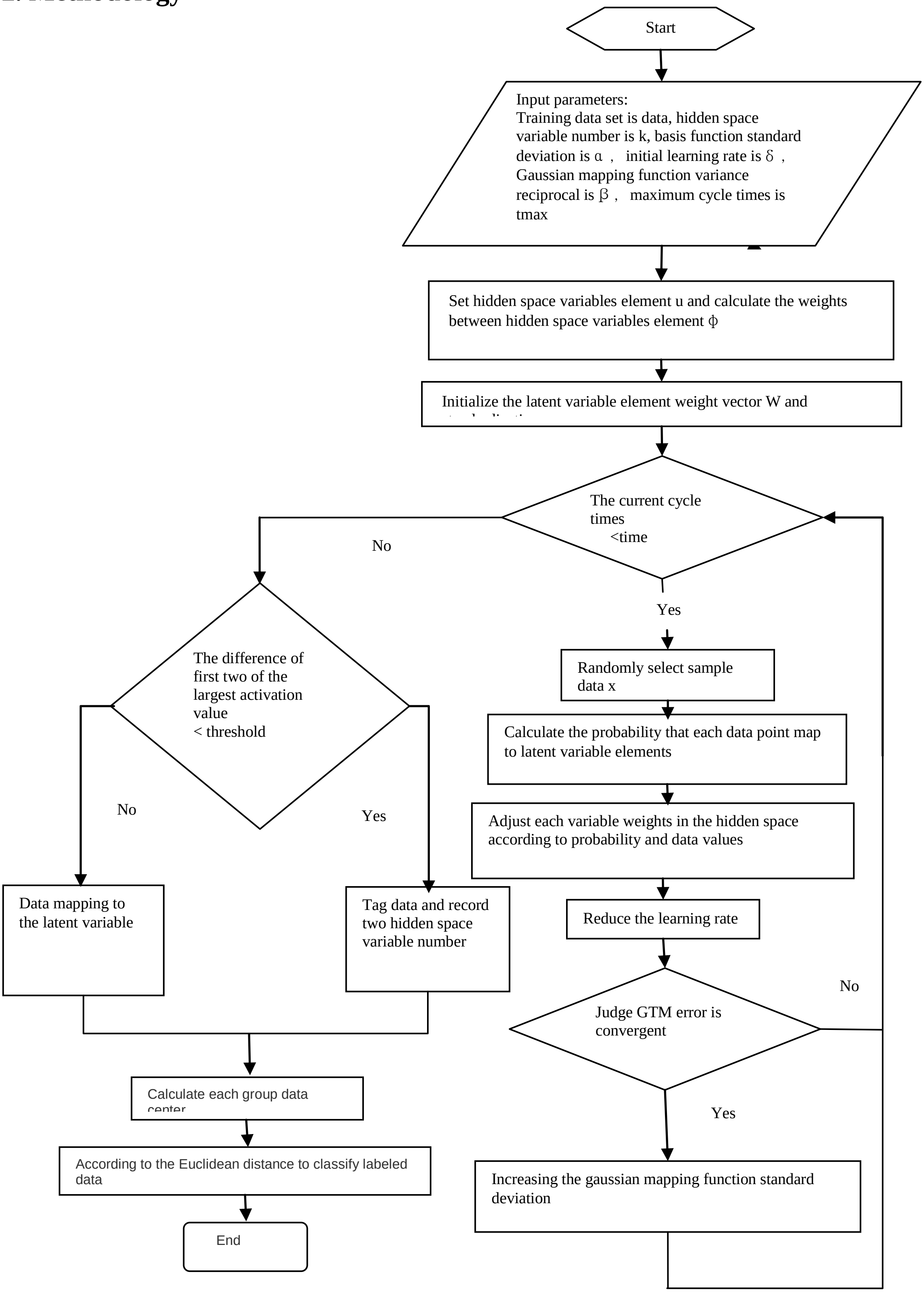


Figure 1. Smap - ED Algorithm Flow

It is similar to GTM - ED model, when Smap -ED decides divide data into groups in the end, it classifies the data by the value of η_i^t, at the moment you can choose two of the largest activated value for comparison, if this difference is within a certain range, data is marked but not classified. After the classification of all datas is completed, according to the center of each group and the group numbers that has two of the largest activated value, then reclassify it. The algorithm process is shown in Figure 1.

The code is that the Smap - ED model is marked the data which cannot be classified, and other data is classified and determined the type. In the code input parameter data means training data set, N means the amount of training data set, K means number of clusters, fi_w means Mapping parameter matrix after training, td means threshold of the biggest difference of the two probabilities, output result typeNo means category number of each data.

```
function typeNo = index_of_closest(data, N, K,fi_w, td)
typeNo = zeros(N,1);
count=0;
for i = 1:N
    lamda_cvector = exp( fi_w * x(i,:)' );
    max_p = max (max(lamda_cvector)); % maximum probability
    [r_1,c_1] = find(lamda_cvector== max_p);% The location of maximum probability
    r=r_1(1);c=c_1(1);
    lamda_cvector(r,c)=0;
    max2_p=max(max(lamda_cvector));% The second large probability
    [r_2,c_2] = find(lamda_cvector == max2_p);% Position
    r2=r_2(1);c2=c_2(1);
    max_12=max_p-max2_p;
    if max_12>td % The difference of two large probability > td,   then normal distribute
category number normally
            typeNo(i) = (c-1) *sqrt(K) + r;
    else
      count=count+1;
      typeNo(i)=0;
      x1=(c-1) *sqrt(K) + r;
      x2=(c2-1) *sqrt(K) + r2;
% Record corresponding category numbers which the number of inputs have two of the
largest probability value

      cluster_num(count,:)=[x1 x2];
   end
end
meanCenter= zeros(K,size(data,2));
for i = 1:K% Circular clustering number
  rnum = find(typeNo == i);
  for ii =1:size(rnum,1)
    meanCenter(i,:)=meanCenter(i,:)+data(rnum(ii),:);
  end
  meanCenter(i,:)=meanCenter(i,:)./size(rnum,1);
end
undefinedNode=find(typeNo == 0); % take out the marked and unclassified data
for i=1:size(undefinedNode,1)  mark1=sum( (data(undefinedNode(i),:)-
meanCenter(cluster_num(i,1),:)).^2 );  mark2=sum( (data(undefinedNode(i),:)-
meanCenter(cluster_num(i,2),:)).^2 );
    if(mark1<mark2)
       typeNo(undefinedNode(i))=cluster_num(i,1);
```

```
    else
        typeNo(undefinedNode(i))=cluster_num(i,2);
    end
end
```

3. Experiment

3.1. The Best Cluster Number for Seismic Attribute Data Set

According to the general situation, select value to input from different model parameter which cluster number is between 4 to 9. According to the running result , using the evaluation indexes algorithm calculates the evaluation results are shown in Table 1 below, in the Table parameter is set successively: number of clusters, maximum cycle times, initial value of learning rate, width of primary function(output unit connection weights gaussian function standard deviation), initial value of Gaussian function mapping variance reciprocal, maximum differential activation value threshold.

Table 1. Smap - ED model Results of Relative Index Table

number of clusters	Parameter setting	**Dunn**	**CH(E+05)**	**I**	**DB^{-1}**
4	4,300,0.8,0.1,1.5,0.2	0.935757	2.350481	0.51501	0.895121
	4,300,0.8,0.01,1.5,0.2	0.933253	2.245722	0.52003	0.803822
	4,300,0.5,0.1,1.5,0.2	0.924541	2.295728	0.52020	0.890331
	4,300,0.5,0.01,1.5,0.2	0.941348	2.405891	0.495022	0.810501
	Mean	0.933725	2.3244555	0.5125655	0.849944
5	5,300,0.8,0.1,1.5,0.2	0.952891	2.315969	0.49287	0.858979
	5,300,0.8,0.01,1.5,0.2	0.951123	2.343766	0.524691	0.897799
	5,300,0.5,0.1,1.5,0.2	0.946221	2.333161	0.513223	0.881232
	5,300,0.5,0.01,1.5,0.2	0.951021	2.235769	0.528322	0.894231
	Mean	0.950314	2.3071663	0.5147765	0.88306
6	6,300,0.8,0.1,1.5,0.2	0.625667	1.997546	0.250511	0.756489
	6,300,0.8,0.01,1.5,0.2	0.656755	2.078434	0.25076	0.850534
	6,300,0.5,0.1,1.5,0.2	0.634656	1.998754	0.250411	0.838676
	6,300,0.5,0.01,1.5,0.2	0.624423	2.039837	0.250283	0.787963
	Mean	0.635375	2.0286428	0.2504913	0.808416
7	7,300,0.8,0.1,1.5,0.2	0.720432	1.768509	0.218084	0.724503
	7,300,0.8,0.01,1.5,0.2	0.402451	1.759184	0.13433	0.770481
	7,300,0.5,0.1,1.5,0.2	0.609541	1.749186	0.11343	0.809362
	7,300,0.5,0.01,1.5,0.2	0.695430	1.798490	0.100249	0.763302
	Mean	0.606964	1.7688423	0.1415233	0.766912
8	8,300,0.8,0.1,1.5,0.2	0.511222	1.748996	0.110852	0.756053
	8,300,0.8,0.01,1.5,0.2	0.529436	1.626572	0.093575	0.768729
	8,300,0.5,0.1,1.5,0.2	0.523121	1.726572	0.103675	0.785059
	8,300,0.5,0.01,1.5,0.2	0.532213	1.591642	0.086117	0.747257
	Mean	0.523998	1.6734455	0.0985548	0.764275
9	9,300,0.8,0.1,1.5,0.2	0.569028	1.514713	0.10739	0.721247
	9,300,0.8,0.01,1.5,0.2	0.562113	1.52321	0.10739	0.762321
	9,300,0.5,0.1,1.5,0.2	0.558122	1.524712	0.117221	0.752417
	9,300,0.5,0.01,1.5,0.2	0.556241	1.403266	0.112818	0.831242
	Mean	0.561376	1.4914753	0.1112048	0.766807

According to Table 1, each number of clusters correspondingly to calculate and get indexes values, according to the formula 2-12 to standardize, and using the WSVF comprehensive evaluation results, the results are shown in the Table 2:

Table 2. Sorting Table of Smap - ED Model Results Evaluation

number of clusters	**Dunn**	**CH(E+05)**	**I**	**DB^{-1}**	**WSVF**
4	0.96108755	1	0.994687928	0.721208142	0.919246
5	1	0.979244106	1	1	0.994811
6	0.261254562	0.644874233	0.365037387	0.371601813	0.410692
7	0.194611509	0.332981484	0.103234634	0.022203842	0.163258
8	0	0.21845686	0	0	0.054614
9	0.087676747	0	0.030392453	0.021317793	0.034847

According to the WSVF value calculated in Table 2, the results are shown in Figure 2 below.

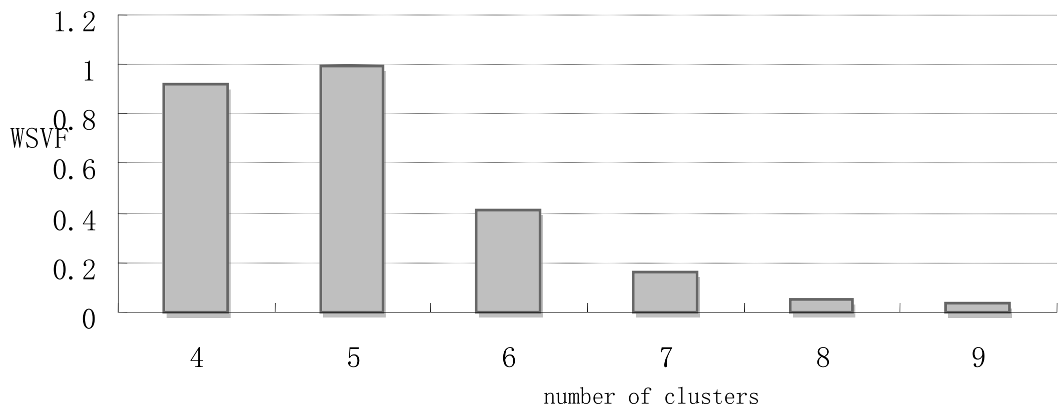


Figure 2. The Values of Smap - ED Clustering Index WSVF

The Figure 2 shows that when the number of clusters is 5, WSVF value is higher, the clustering results of this model for seismic attribute data is better. So, determining the Smap - ED reservoir identification model algorithm's best cluster results for the data set is 5. By observing the visual result Figure which gets from model that inputs different clustering number , we can also observe which is better roughly, by comparing the Figure 3, Figure 4 and Figure 5, according to the known geological information, it can be seen in the case of clustering number is 5 model, clustering effect is better. Through the visual model clustering results graph, we can be more intuitive and easily to identify the unknown reservoir.

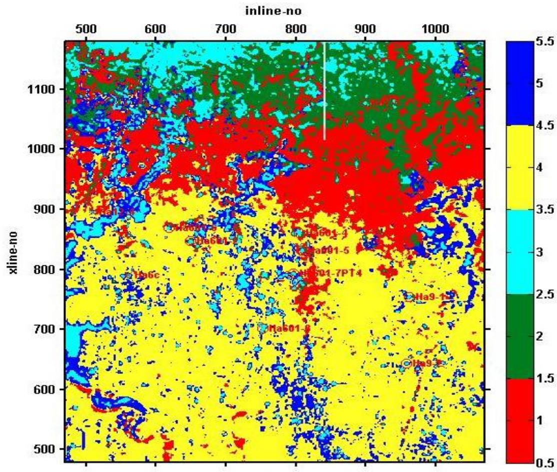


Figure 3. The Result of 5 Kinds of Smap-ED Model Result

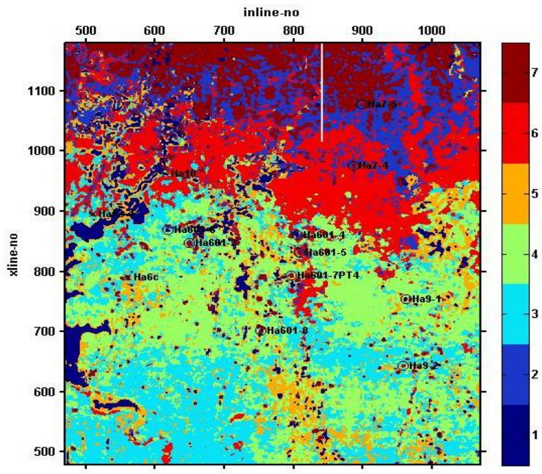


Figure 4. The Result of 7 Kinds of Smap-ED Model Result

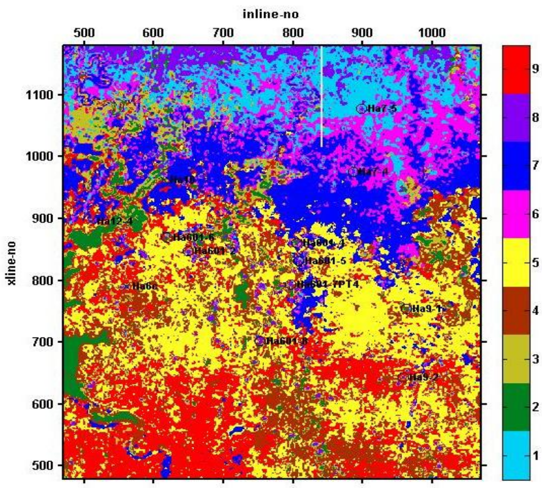


Figure 5. The Result of 9 Kinds of Smap-ED Model Result

3.2. An External Evaluation for The Rationality of the Model Results

According to the 5% predetermined known label information, we can get the calculation results of external evaluation index Rand and F - Measure by model results, as shown in Table 3.

Table 3. Smap - ED Model Results of External Index Table

Parameters	Rand	F-Measure
5,300,0.8,0.01,1.5,0.2	0.881122	0.125472

4. Comparison

The results from the models including Smap-ED, S-Map, SOM, GTM and GTM-ED are compared and analyzed.

4.1. Comparative Analysis of Relative Index

Relative indexes evaluate clustering results by the degree of separation in different classes and the tightness in a class, by comparing the model results of relative index, can reflect the result of model for data clustering is good or bad. Comparison results as shown in Figure 6.

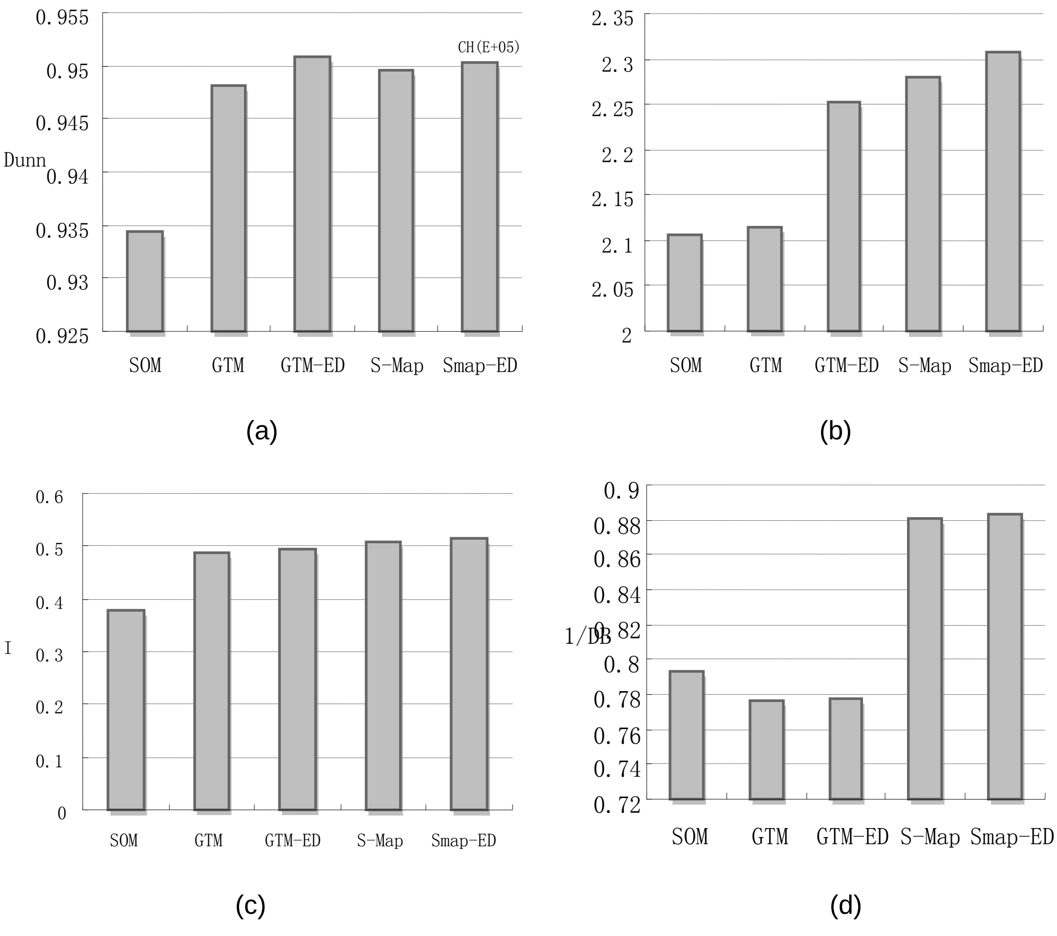


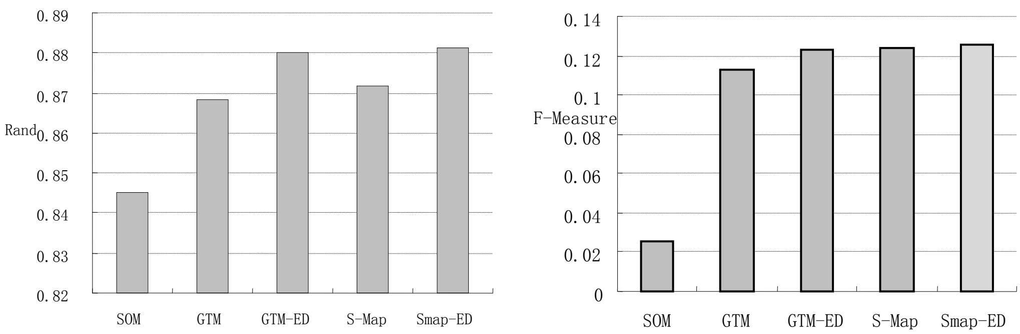
Figure 6. Comparison Diagram of Model Result Relative Index

According to the results of the relative evaluation indexes, Smap - ED model clustering results are better, compared with the original S - Map effect does have some improvement, but the improvement is not particularly outstanding.

4.2. The Comparative Analysis of External Indexes

The size of the external appraisal index directly reflects the accuracy of clustering results, the comparison of the two model clustering results external parameter values are shown in Figure 7.

The external indexes results show that the accuracy of Smap– Ed model is better compared to the model S - Map. To some extent, the accuracy and rationality of this clustering results in the 5 models is the best.



(a) (b)

Figure 7. Comparison Diagram of Model Result External Indexes

4.3. The Comparative Analysis of Running Time

By counting running time, we can get the comparative analysis of the running efficiency for the five models algorithm. The comparisons of running time are shown in Table 4 and Figure 8.

Table 4. Comparison Diagram of Model Running Time

Running time	1	2	3	4	5	6	7	8	9	10
SOM	21.51	21.52	20.91	20.96	20.77	20.58	20.52	20.41	21.08	21.11
GTM	56.13	55.1	57.29	54.38	54.45	54.7	53.26	53.69	55.04	54.89
GTM-ED	54.6	55.26	55.44	55.51	56.87	56.27	57.3	56.72	58.22	56.77
S-Map	21.98	23.2	22.32	22.14	23.21	22.88	23.36	23.54	24.17	22.91
Smap-ED	22.82	23.71	25.9	21.71	23.4	23.47	22.79	24.19	24.51	24.23

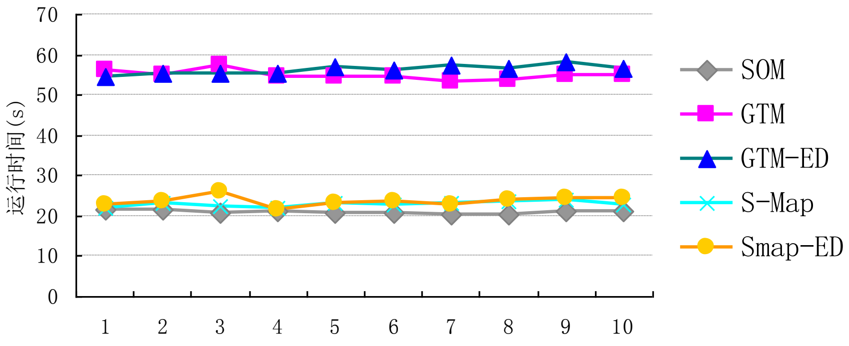


Figure 8. Comparison Diagram of Model Running Time

It can be seen from the model algorithm average running time, the running time of running time SOM, S - Map and Smap - ED algorithm are close, all smaller than the running time of GTM and GTM - ED obviously, thus we can know algorithm efficiency comparison.

Through the above statistical analysis of evaluation indexes, we can get the following conclusion: from effectiveness evaluation index, compared with SOM, GTM clustering effect is better, for GTM-ED model, from the index values we can see it can improve clustering effect. For the S-Map model, it is more successful in the application to seismic attribute data set, because no matter look from the result of the evaluation indexes, or from the running time of the algorithm, S - Map algorithm has advantages all the time. Due to the S - the basic steps are following SOM algorithm, but using the GTM research probability and expectation-maximization to update the weight vector and the parameters, so the algorithm can achieve the similar levels of clustering of GTM by a small amount of time, and in many cases the S-Map clustering effect is better than GTM, that's because the S–Map algorithm combines with SOM self-organizing, so it reduces the sensitivity of parameter selection. The Smap - ED model also have effect of certain improvement.

5. Conclusion

This paper introduces the research and application of reservoir identification model based on Smap - ED. First it discusses the thought of model, and describes the process and implementation. Then it is the experimental analysis, the relative indexes use to determine the model and the optimal number of clusters for seismic attributes data set, and make a comparative analysis from the relative index, external index and running time of SOM, GTM, GTM - ED, S - Map and Smap - ED, the conclusion is: the Smap - ED model of seismic attribute data clustering effect is slightly better than the original model S - Map, And the running time of the algorithm and SOM is close. While this article has made some achievements, but research on the application of data mining to seismic attribute data analysis is still in its initial stage, there are a lot of problems need further research.

Acknowledgments

This work has been supported by Fujian Province Education Department of China under Grant Nos. JB12209, the Research Program of RIPED (No. 101002kt0b52135), the Scientific Research Foundation for the Returned Overseas Chinese Scholars, the Natural Science Foundation of Fujian Province of China (No. 2015J05129), and the National Natural Science Foundation of China (No. 61502402).

References

[1] W. Y. Syrmosaurus and C. M. Zhang, "Oil and gas reservoir geology", Beijing: petroleum industry press, **(1996)**.
[2] L. Jianguo, "Improved K - means algorithm and its application in oil production data analysis", Beijing university of posts and telecommunications, **(2010)**.
[3] A. Roy, A. S. Romero-Peláez, T. J. Kwiatkowski and Kurt J., "Marfurt Generative topographic mapping for seismic facies estimation of a carbonate wash, Veracruz Basin, southern Mexico", Interpretation, vol. 2, no. 1, **(2014)**, pp. 31-47.
[4] Maulik U. and Bandyopadhyay S., "Performance Evaluation of Some Clustering Algorithms and Validity Indices", IEEE PAMI, vol. 24, no. 12, **(2002)**, pp. 1650-1654.
[5] Y. H. Zhao and R. Y. Peng, "Seismic attributes of cloud rocks reservoir identification study", Journal of engineering geophysics, vol. 10, no. 6, **(2013)**, pp. 891-895.
[6] L. Weiping, H. Q. Li and S. Ning, "Data mining application in seismic attribute analysis", Journal of geophysics in China, **(2012)**, pp. 509-512.
[7] A. R Brown, "Seismic attribute and classification", Journal of foreign oil and gas exploration, **(1997)**, pp. 529-530.
[8] Q. Chen and S. Sidney, "Used for reservoir prediction and monitoring of the seismic attribute technology", Journal of foreign oil and gas exploration, vol. 10, no. 2, **(1998)**, pp. 220-231.
[9] Y. Ruichao, "DBSCAN algorithm in the application of seismic facies classification", Xi an university of science and technology, **(2011)**.
[10] D. J. Chardonnay, "K - means algorithm is improved and its application in geological exploration seismic data analysis", Beijing university of posts and telecommunications, **(2012)**.
[11] Theodoridis S. and Koutroubas K., "Pattern Recognition", Academic Press, **(1999)**.
[12] Yang, sorer J a. data compression, "Mohame. Clustering validity evaluation review", Computer application research, vol. 25, no. 6, **(2008)**, pp. 1630-1632.
[13] D. ell, "Separated Clusters and Optimal Fuzzy Partitions", J Cybern, vol. 4, no. 1, **(1974)**, pp. 95-104.
[14] Calinski T. and Harabasz J., "A Dendrite Method for Cluster Analysis", Journal of Comm. in Statistics, vol. 3, no. 1, **(1974)**, pp. 1-27.
[15] Davies D. L. and Bouldin D. W., "A Cluster Separation Measure", IEEE the Transactions on Pattern Analysis and Machine Intelligence, vol. 1, no. 2, **(1979)**, pp. 224-227.
[16] L. Y. Xue-Dong Gao and W. Sen, "The combination of clustering validity evaluation method", Computer engineering and application, vol. 47, no. 19, **(2011)**, pp. 15-17.
[17] Z. Yong, "Based on the combination of the multi-objective decision clustering model and its application", University of electronic science and technology, **(2013)**.
[18] Dimitriadou E., Dolnicar S. and Weingessel A., "An Examination of the Indexes for Determining the Number of Cluster in Binary Data Sets", J Psychometrika, **(2002)**, pp. 137-160.
[19] T. K ohonen, E. O Ja, G. Imula, Dr. Isa and J. K. Angas, "The Engineering Applications of the Self - Organizing Map", Proceedings of the IEEE, **(1996)**, pp. 1358-1384.
[20] Svensen M., "GTM research: The Generative Topographic Mapping", Tianjin University, **(1998)**.
[21] K. Kiviluoto and E. Oja, "S-Map: A Network with a Simple Self-Organization Algorithm for Generative Topographic Mappings", In Advances Neural Information Processing Systems, **(1998)**, pp. 549-555.
[22] R. T. Proyer, "Development and initial assessment of a short measure for adult playfulness: The SMAP", Personality and Individual Differences, vol. 53, no. 8, **(2012)**, pp. 989-994.
[23] X. Wu, J. P. Walkera, N. N. Dasb, R. Pancierac and C. Rüdigera, "Evaluation of the SMAP brightness temperature downscaling algorithm using active–passive microwave observations", Remote Sensing of Environment, vol. 154 no. 12, **(2014)**, pp. 210-221.
[24] M. Miernecki, J. P. Wigneron, E. L. Baeza, Y. Kerr, R. D. Jeu, G. J. M. De Lannoy, T. J. Jackson, P. E. O'Neillf, M. Schwank, R. F. Moran, S. Bircher, H. Lawrenc, A. Mialon, A. Al Bitar and P. Richaume, "Comparison of SMOS and SMAP soil moisture retrieval approaches using tower-based radiometer data over a vineyard field", Remote Sensing of Environment, vol. 154, no. 11, **(2014)**, pp. 89-101.
[25] R. T. Proyer, "Involvement of SMAp in the intention-related long latency stretch reflex modulation: A TMS study", Neuroscience, vol. 246, no. 8, **(2013)**, pp. 329-341.
[26] Dr. Kurt J., "Marfurt.Latent Space Classification of Seismic Facies", the University of Oklahoma, **(2013)**.

