doi:10.6043/j.issn.0438-0479.201611020

**自组织映射神经模型改进及其在储层预测中的应用**

鲍彬彬，吴清强\*

（厦门大学软件学院，福建厦门361005）

**摘要：**储层是有一定的连通孔隙，能存储和渗滤自组织映射神经油气的岩层。但并非所有储层都存储着油气。将数据挖掘技术应用于石油勘探开发中，建立准确的储层预测模型。自组织映射神经（Self-organization Map，SOM）网络是一种具有广泛应用的无监督聚类算法。该聚类算法可用于储层的预测。类内相似度是度量聚类结果质量常用的评价指标。类内相似度越大，聚类结果越好。将类内相似度这一评价指标引入SOM神经网络学习过程中，通过调整类内最小相似度来指导SOM神经网络学习，使得平均类内最小相似度最大，提高SOM神经网络聚类结果质量。改进的SOM算法解决了部分神经元过度利用和欠利用的问题。实验表明，改进后的SOM算法聚类结果确实得到了提高。

**关键词：**自组织映射神经；类内最小相似度；改进

**中图分类号：**TP391 **文献标志码：**A

分析地震属性数据是希望从这些数据能得到有关岩性和储层物性的信息。这些信息帮助石油勘探工作者更好确定油田的位置，降低开发失败的可能性，节省开发费用，加强地震数据在油田开发领域的应用。

随着石油勘探开发的不断深入，想要对海量的地震属性数据进行归类、标定，有必要将数据挖掘技术应用于石油勘探和开发中。20世纪80年代以来，数据挖掘技术在储层预测领域中得到了广泛的应用。世界各国学者提出了灰度关联度分析法、线性和非线性分类、模糊判别技术等储层预测方法。灰度关联分析是系统分析中比较简单、可靠的一种方法，对样本的数量没有多大要求，样本的分布也不需要是常规典型的。这个分析法最大缺陷是要设置各项指标的最优值。线性和非线性分类这类储层预测法比较成熟，抗噪能力强，但是随着石油勘探开发深度和广度的不断扩展，非常规油气藏逐渐成为主要勘探对象，对于海量复杂待预测地震数据，用少量有标签数据难以建立高准确的分类器。模糊判别技术使用软划分概念，使储层预测更符合实际，但难以建立准确的隶属度函数[1]。

神经网络是解决复杂非线性映射问题的有效手段，有较强容错能力。BP神经网络及其与退火算法、遗传算法组合的神经网络常用于储层预测[2]。SOM神经网络是数据挖掘中的一种无监督聚类算法。该神经网络不需要已知的类别信息，能自动对输入模式进行聚类。SOM聚类算法采用WTA（Winner Take All）竞争机制，对噪音数据不敏感。并且此算法有从高维数据映射到低维时数据保持拓扑结构的特性，竞争层神经元经过训练后权重分布于输入模型的空间分布趋于一致[3]。

SOM聚类算法也存在着很多不足的地方，会影响到SOM神经网络的性能，降低聚类结果的质量[4]，在此基础上对地震属性数据进行分析，获取的可能是不正确的岩性和储层物性的信息。本文针对SOM神经网络训练过程中部分神经元会过度利用和欠利用的问题，引入类内相似度聚类评价指标来指导SOM神经网的训练，使竞争层每个神经元得到充分利用，提高聚类结果质量。

## 1 SOM模型基本简介

SOM通常包括竞争环节，合作环节和自适应环节这三个环节。在竞争环节中，根据WTA竞争学习规则，计算每个神经元与输入向量的欧式距离，与输入向量距离最短的神经元作为竞争胜利者。在合作环节中，竞争胜利元和在以竞争胜利神经为中心的领域内的神经元都按照一定的学习规则被激励，与竞争胜利神经元邻近的神经元被激励程度大。在自适应环节中，适当调整竞争胜利神经元领域内所有神经元的权值，使的竞争层神经元的权向量随输入向量进行改变，神经元对之后相似的输入向量的有一个加强的响应[5]。

SOM神经网络采用的是WTA竞争学习规则，训练神经网络时，每次输出层只有一个最强神经元获胜，该神经元输出为1，其他神经元输出为0。这会造成以下情况：如果有些输出层神经元与输入模式相差甚远，则这些神经元在训练过程中始终不能获胜，成为“死神经元”；相反有些神经元获胜次数过多，出现神经元过度利用的问题。为了解决神经元神经元欠利用和过度利用的问题，提出了SOM-CV、SOM-C、ESOM、TASOM、DSOM等代表性改进算法。这些算法对SOM竞争学习规则提出各自改进方法，例如SOM-C算法给输出层神经元设置一个阈值减少经常获胜神经元获胜机会[6]，在一定程度上避免了部分神经元欠利用和过度利用的问题，聚类结果质量得到了提高。本文借鉴于Ai-xiang Sun提出的基于等离差理论的一种学习竞争规则改进算法。该算法以平均类内离差最小即平均类内相似度最大目标，引入类内相似度，指导神经网络的训练，改进了竞争学习规则[7]。不同的是本文使用最小类内相似度来表示类内离差。

## 2 SOM算法的改进

在储层预测中，通过SOM神经网络建立地震属性数据与储层类别之间的聚类关系。地震属性数据是地震波经过数据变换推导出的，携带有地下岩层信息[8]。SOM神经网络训练过程会发生大量地震属性数据聚集于部分神经元，相应的另一部分神经元较少被激励问题。本文引入等离差理论，改进SOM神经网络的竞争学习规则来解决部分神经元欠利用和过度利用的问题。

聚类是根据某种相似度度量将相似的数据归为一类，不相似的数据分到不同类别中。相似度度量方法主要有相关系数法和空间距离法[9]。SOM神经网络就是通过欧式距离度量相似度。在实际应用中，要先对一个聚类结果进行评价，根据评价结果决定是否采用这次聚类结果。评价聚类结果好坏的最直观想法是类内数据越相似越好，类间数据越不相似越好。类内相似度是评价聚类结果质量的一个重要指标。直接将类内相似度评价指标引入SOM神经网络训练过程中，以类内相似度最大为目标，训练神经网络，这样可以获得更好的聚类结果质量。

SOM算法学习过程的竞争环节中，计算输入向量与竞争层中每个神经元的欧式距离，以此为依据，将与输入向量最近的神经元定为竞争胜利神经元。该输入向量就归属于这个竞争胜利神经元所代表的类别。此欧式距离用来于度量输入向量跟类别的相似度。类别中存在着一个相似度最小即欧式距离最大的样本。这一样本会影响SOM算法聚类结果质量[10]。类内最小相似度越大，表明这个类越紧密，类内的样本数据越相似。因此在SOM算法竞争环节中把类内最小相似度考虑进去。SOM神经网络经过反复训练之后，使得平均类内最小相似度最大。定义新的输入向量与竞争层中每个神经元的距离公式：



其中X表示输入向量，Wi表示第i个神经元的权向量，Si表示第i个类别空间，表示第i个类别空间的最小相似度。

SOM改进算法步骤如下：

1. 用小随机数初始化竞争层神经元的权向量，并将每个神经元节点的类内最小相似度初始化为1，给定学习率和领域半径初始值。
2. 从样本空间中随机选取一个样本输入到网络中。
3. 用新的距离公式确定竞争胜利神经元。
4. 更新竞争胜利神经元和以其为中心的领域内所有神经元的权向量，调整学习率和领域半径。
5. 更新竞争层每个神经元的类内最小相似度。若竞争层神经元结点对应的类别样本数非零，则按照下面公式更新此神经元的类内最小相似度。



式中num1是对应的类别样本数非空的神经元个数。若竞争层的神经元结点对应的类别样本数为零，则按下面公式更新类内最小相似度[11]。



式中num2是对应的类别样本数为零的神经元个数。

1. 返回步骤(2)，直到达到最大迭代次数为止。

## 3 实验结果及分析

SOM神经网络的竞争层是由一维或者二维神经元组成。研究表明，一维SOM聚类性能并不比二维SOM差，而且，与二维SOM相比，一维SOM更容易识别类边界，更清楚表达样本数据的相似度和类之间的相邻关系[9,12]。因此本实验采用一维SOM神经网络。

聚类结果有效性关系到它所起的作用大小。常用的聚类有效性评价方法可分为外部评价法、内部评价法和相对评价法。外部和内部评价法是通过判断一个数据集与预先已知结构是否具有较好的相符程度来衡量聚类结果的有效性。这两大类评价方法都是基于统计测试，计算复杂度比较高[13-14]。相对评价法则是测试聚类算法不同的假设和参数，寻得最优聚类结果[14]。本文运用数据可视化技术对聚类结果进行可视化，可以通过人的视觉观察来评价聚类结果的质量。

本实验使用相对评价法中的几个常用聚类评价指标：Calinski-Harabasz指标(CH)，Dunn指标，Davies-Bouldin指标(DB)。CH指标和Dunn指标都是分离度和紧密度的比值。CH指标用类中的点到类中心的距离的平方和来表示类的紧密度，用类中心到数据集中心点的距离来表示数据集的分离度。Dunn指标用类与类之间最小两点距离来度量类的分离度，用类中最大两点距离来度量类的紧密度。DB指标是类内分离度与类间分离度的比值[15]。DB指标是越小，聚类结果越好而CH指标，Dunn指标是越大，聚类结果越好。为了与CH指标，Dunn指标保持一致，使用DB指标的倒数。

实验环境：window 7， Intel Core i7 四核处理器，主频 2.00GHz，8 GB内存。

本实验参数设置：竞争层神经元分布为7个神经元的一维阵列，初始学习率为0.9，初始领域半径为3。

基本SOM算法和改进之后的算法使用塔里木盆地塔北哈6工区地震属性数据进行实验，该工区是具有典型的“串珠状反射”特征的碳酸盐储层，独有的地震地质现象，如图1所示，此图为该工区的一条地震剖面。使用的样本数据包含13个地震属性：10HZ分频数据体，20HZ分频数据体，30HZ分频数据体，15-25度叠加数据，25-35度叠加数据，5-15度叠加数据，15-40HZ分频振幅差，叠前反演纵波速度，叠前反演横波速度，叠前反演纵横波速度比，叠前反演纵波阻抗，弹性阻抗，叠前反演密度体。其中20HZ分频数据体属性，叠前反演纵横波速度比属性，25-35度叠加数据属性，15-40HZ分频振幅差属性可视化如图2-5所示。

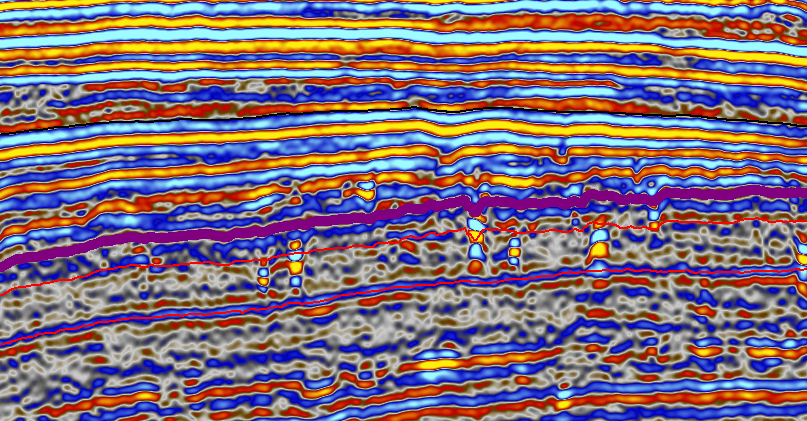


图1地震体剖面

Fig.1Seismic section

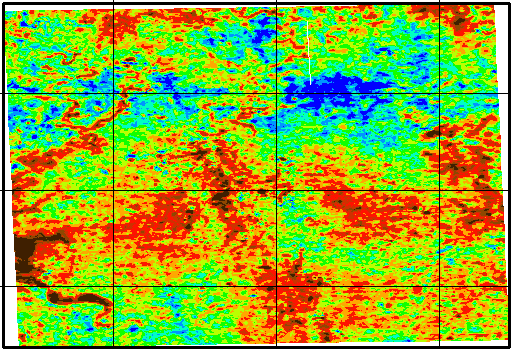
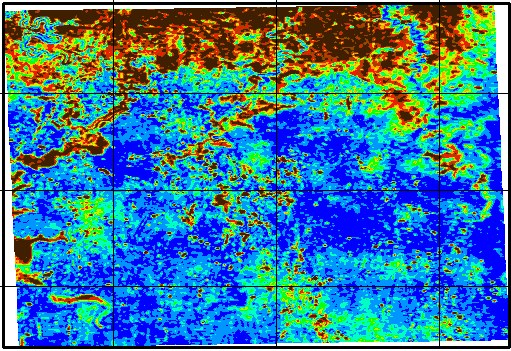


图2 20HZ分频数据体 图3叠前反演纵横波速度比

Fig.2 20HZ frequency data Fig.3 pre-stack inversion P-wave-to-S-wave velocity ratio

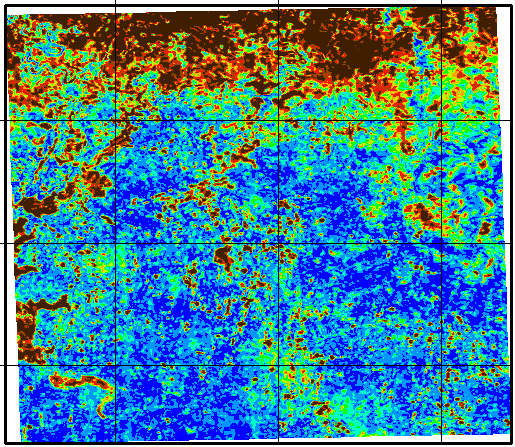
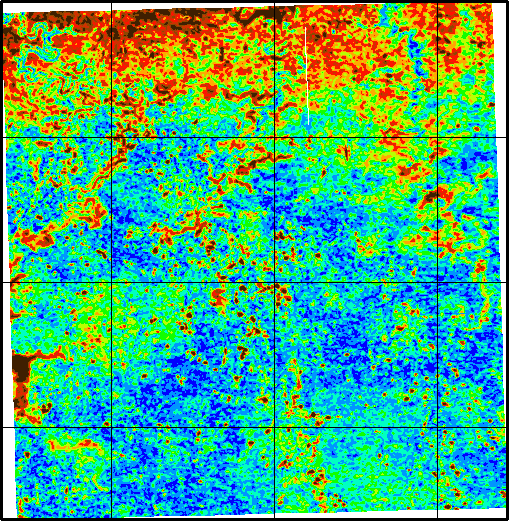


图4 25-35度叠加数据 图515-40HZ分频振幅差

Fig.4 25-35 degree superposed data Fig.5 15-40HZ frequency amplitude difference

样本数据经过标准化处理。为了避免随机性的影响，统计每个聚类评价指标的10次结果，取平均值。实验结果如表1，图6-7所示。

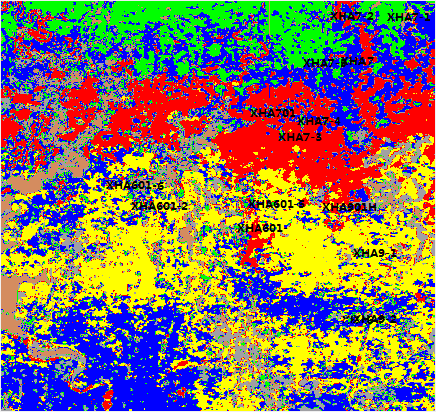


图6基本SOM聚类结果可视化图

Fig.6 basic SOM clustering result visualization

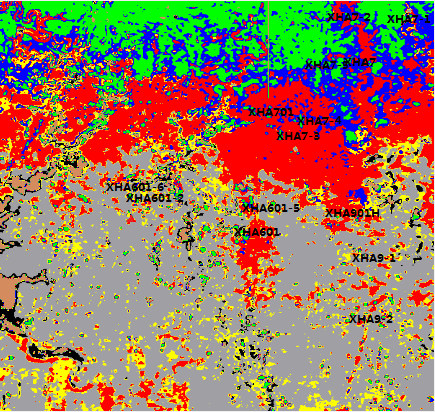


图7改进后的SOM聚类结果可视化图

Fig.7improved SOM clustering result visualization

表1基本SOM和改进SOM聚类结果评价指标对比

Tab.1 Comparison of Evaluation Indexes of Cluster Result of Basic SOM and Improved SOM

|  |  |  |  |
| --- | --- | --- | --- |
| 算法 | CH指标X105 | Dunn指标 | DB-1指标 |
| 基本SOM | 1.6038 | 0.1739 | 0.6932 |
| 改进SOM | 1.7981 | 0.3085 | 1.1249 |

图6，图7分别是基本SOM神经网络和改进SOM神经网络聚类结果二维可视化。

本实验两个算法是将地震数据分为7类，在可视化图中分布用7种颜色表示。图中的“X”表示的是实际已开发的油井。两个算法聚类可视化图都呈现出哈6工区的主要地质特征：上部分绿色，蓝色和红色三个颜色混合的地区是带状熔岩，中上部分红色区是富油区，多个油井落入该区域。中下部分是串珠储层，左下部分褐色区是古河道。对比基本SOM聚类结果图，改进SOM算法结果图中有明显的黑色类别区，对古河道的刻画更加清晰，灰色类别区面积增加较大。改进SOM引入等离差原理，以总类内离差最小为目标指导神经网络的训练，一定程度上解决了神经元欠利用和过度利用的问题，聚类结果图中类边界更为清晰，串珠储层更为明显，更多油井落在红色表示的富油区，对石油勘探研究人员分析和判断储层位置更为有用。从表1中，可以看出：对于相同的样本集，改进后的SOM算法三个指标均优于基本SOM算法。改进后的SOM算法CH指标比基本SOM算法的高12.1%，Dunn指标大77.4%，DB-1指标要高62.3%。这个说明了改进后的SOM算法聚类结果质量更好。将类内最小相似度引入SOM算法学习过程这一改进提高了SOM算法的性能。

## 4 结 论

本文提出一种SOM算法改进方法，在SOM算法竞争环节中引入类内最小相似度，网络学习过程中不断调整每个神经元的类内最小相似度，使得平均类内最小相似度最大，来提高聚类结果质量，并通过实验验证。实验结果表明改进后的聚类结果质量更高。

本文SOM神经网络竞争层神经元采用一维分布，结构简单固定。在未来工作中，将对竞争层神经元结构进行研究，提高聚类结果质量。

**参考文献:**

[1] 刘正锋, 燕军. 模糊识别方法在储层识别中的应用[J]. 西南石油大学学报自然科学版, 1998, 20(3):4-6.

[2]刁凤琴, 诸克君, 於世为. 一种优化的BP神经网络算法在石油储层预测中的应用[J]. 系统管理学报, 2008, 17(5):499-503.

[3] 杨黎刚，苏宏业，张英. 基于SOM聚类的数据挖掘方法及其应用研究[J]. 计算机工程与科学，2007, 29（8）:133-136.

[4] Li Y, Pan F. Application of Improved SOM Neural Network in Manufacturing Process Quality Control[C]//Atlantis Press. Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering. Paris: Atlantis Press, 2013:1154-1157.

[5] De A, Guo C. A vector quantization approach for image segmentation based on SOM neural network[C]// Springer-Verlag. International Conference on Advances in Neural Networks. Berlin, Heidelberg: Springer-Verlag, 2013:612-619.

[6]杨占华, 杨燕. SOM神经网络算法的研究与进展[J]. 计算机工程, 2006, 32(16):201-202.

[7] Sun A X. Improved SOM Algorithm-HDSOM Applied in Text Clustering[C]//IEEE. 2010 International Conference on Multimedia Information Networking and Security. Piscataway, NJ: IEEE, 2010:V6-42 - V6-44.

[8] 杨培杰, 印兴耀, 张广智. 模糊C均值地震属性聚类分析[J]. 石油地球物理勘探, 2007, 42(3):322-324.

[9] 陈善学, 杜峰, 吴立彬. 一种改进的等误差自组织特征映射矢量量化算法[J]. 重庆邮电大学学报(自然科学版), 2011(02)：155-157.

[10] Liu F, Zhang G L, Sun Y J, et al. Mapping the Three-Dimensional Distribution of Soil Organic Matter across a Subtropical Hilly Landscape[J]. Soil Science Society of America Journal, 2013, 77(4):1241-1253.

[11] Akinduko A A, Mirkes E M, Gorban A N. SOM: Stochastic initialization versus principal components[J]. Information Sciences, 2015, 364–365:213-221.

[12] 于鷃. 基于一维SOM神经网络的聚类及数据分析方法研究[D]. 天津：天津大学, 2009: 1.

[13] Turkay C, Parulek J, Reuter N, et al. Integrating cluster formation and cluster evaluation in interactive visual analysis[C]//ACM. Spring Conference on Computer Graphics. New York, NY: ACM, 2011: 77-86.

[14] 杨燕，靳蕃，KAMEL M. 聚类有效性评价综述[J]. 计算机应用研究, 2008, 25(6):1630-1632.

[15] 刘燕驰, 高学东, 国宏伟, 武森. 聚类有效性的组合评价方法[J]. 计算机工程与应用, 2011(19):15-16.

**Improvement and Application in Reservoir Prediction of Self-organizing Map Model**

BAO Binbin, WU Qingqiang\*

(Software School, Xiamen University, Xiamen 361005, China)

**Abstract:** Reservoir is rock strata which can store and filtrate oil gas with some connected pores. Data mining is applied to petroleum exploration and development to establish accurate reservoir prediction model.SOM (Self-organizing Map) network is an unsupervised clustering algorithm with wide application. The clustering algorithm can be used to predict the reservoir. Intra-Class similarity degree is a commonly used evaluation index to evaluate the quality of the clustering results. Intra-Class similarity degree can be used to weigh the cluster result. We introduce intra-class similarity degree into the process of SOM neural network learning, adjust intra-class minimum similarity degree into the process of SOM neural network learning, adjust intra-class minimum similarity degree to guide SOM neural network learning, which makes the average intra-class similarity degree maximum and improves the quality of cluster result. The improved SOM algorithm can solve the problem of excessive use and less use of some neurons. The experiment shows that the improved SOM algorithm has improved the clustering results.

**Key word:** self-organizing map; intra-class minimum similarity degree; improvement