

一般化的灰 SOM 模型及其性能评估

潘志松 陈松灿 张道强

(南京航空航天大学信息科学与技术学院 南京 210016)

摘 要 自组织特征映射是 Kohonen 提出的一种人工神经网络模型,它能把高维的信息数据以有序方式映射到低维空间上,形成一种拓扑意义上的有序图.但其权值的调整仅考虑了学习率及输入模式与获胜神经元及其邻域,却忽略了输入模式分量与全体参与竞争的神经元权值向量间的某种相关关系. Hu 等人利用灰关系系数(GRC)的描述方法显式地刻画这种关系.尽管如此,但 GRC 仍忽略了输入模式与所有参与竞争的神经元权值间的整体描述关系.该文通过在 GRC 中引入该整体关系,提出了一个一般化的灰关系模型 G^2 SOM,并将 GSOM 作为其特例,进而侧重就三种特殊的函数关系进行了性能评估.最后,通过实验验证了 G^2 SOM 的有效性与可行性.

关键词 灰关系系数;自组织特征映射;分类;一般化
中图法分类号 TP18

Generalized Grey SOM and Their Performance Evaluations

PAN Zhi-Song CHEN Song-Can ZHANG Dao-Qiang

(College of Information Science and Technology, Nanjing University of Aeronautics & Astronautics, Nanjing 210016)

Abstract The Kohonen self-organizing maps is a typical artificial neural networks model and algorithm that implements a nonlinear feature projection from the high-dimensional space of signal data into a low-dimensional array of neurons in an orderly fashion. The mapping tends to preserve the topological relationships of signal domains. But its weight adjustment is determined only by its learning rate and the difference between the input pattern and the winner neuron's weight. It seems that the SOM obviously ignores some (implicit) correlation relationships during the learning, which actually exist between the input pattern and the weights of all the nodes that participate in competition. Grey relational coefficients (GRC), which characterize and stress the aforementioned correlation relationships, are explicitly introduced into the learning rule of the traditional SOM. By more analyses, the so-defined GRCs still ignore some whole measure relationship as a similarity between the input pattern and weights. In addition to the whole characteristic to the original GRCs, The SOM is generalized by the designed three GRC functions, namely G^2 SOM. (Generalized Grey Self-organizing Maps). Finally, the experiments on the benchmark datasets confirm their validities and feasibilities over the G^2 SOM in this paper.

Keywords grey relation coefficients; self-organizing maps; classification; generalization

1 引 言

Kohonen 提出的自组织特征映射理论(Self-

Organizing Maps, SOM)^[1]具有聚类、自组织、自学习以及可视化的功能,已广泛地应用于模式识别、故障诊断等领域.但因其学习规则在对竞争获胜的神经元及其邻域的权值进行调整时,仅考虑了学习率

收稿日期:2002-11-07;修改稿收到日期:2003-07-10.本课题部分得到国家自然科学基金(60271017)、江苏省自然科学基金(BK2002092)、江苏省“青蓝工程”基金和留学回国人员基金资助.潘志松,男,1973年生,博士研究生,主要研究方向为信息系统安全、入侵检测系统、神经网络. E-mail: Idsystem@sina.com. 陈松灿,男,1962年生,教授,博士生导师,主要研究方向为神经网络、模式识别. 张道强,男,1978年生,博士研究生,主要研究方向为神经网络、模式识别.

以及当前输入模式与权值之间的差异,却忽略了在输入模式与所有参与竞争的神经元权值之间实际存在的某种隐关系. Deng 的灰色系统理论^[2]通过对子系统间的灰关系分析,能很好地处理有限的、不完全的输出数据序列. 他们认为现实世界里的任意两个子系统之间都存在某种隐含关系. 而学习规则中输入模式与所有参与竞争的神经元的权值也可视作两个子系统,因此文献^[3]将灰关系引入了 SOM 中,构建了 GSOM (Grey SOM), 实验表明和 SOM 相比, GSOM 有更好的分类和推广性能.

尽管如此, GSOM 主要考虑了各分量间的局部关系,忽略了输入模式与所有参与竞争的神经元权值间的整体描述关系. 本文以此为基础在保留了原灰关系的同时,引入整体特性,从而兼顾了两者,建立了一个局部和总体相融合的灰关系,进一步设计了一般性的 G^2RC (Generalized Grey Relation Coefficient), 并侧重就指数、多项式和正切的三种灰关系函数进行讨论; 将 G^2RC 和 SOM 结合, 并采用投票表决策算法形成 G^2SOM 分类模型, 最终通过实验对灰关系系数的三种函数形式及其效果进行了性能评估.

2 SOM 模型

SOM 网络由两层构成, 第一层是输入层, 负责数据的输入, 第二层是 Kohonen 层, 按二维形式排成一个节点矩阵, 输入节点处于其下方, 如图 1 所示.

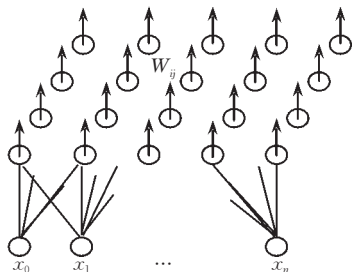


图 1 SOM 网络结构

具体的自组织特征映射算法^[4]如下.

1. $IT=0$, 对所有从输入节点到输出节点的连接权值赋以小的随机数, 选择学习率 $\alpha(t)$ 和 $NE_{i^*}(t)$ 随时间变化的形式, M 为神经元个数.

2. 对网络输入模式 $\mathbf{X}=[x_1, x_2, \dots, x_n]^T$.

3. 计算输入与所有输出节点连接权矢量的距离:

$$d_i = \sum_{j=1}^n (x_j - w_{ij})^2, \quad i = 1, 2, \dots, M \quad (1)$$

4. 选择具有最小距离的节点 i^* 作为获胜节点:

$$d_{i^*} = \min_i \{d_i\} \quad (2)$$

5. 调整输出节点及其几何邻域内的节点所连接的权值矢量:

$$w_i(t+1) = w_i(t) + \alpha(t)[\mathbf{X} - w_i(t)], \quad i \in NE_{i^*}(t) \quad (3)$$

6. 若还有训练样本数据, 则 $IT=IT+1$, 转 2; 否则, 转 7.

7. 训练结束.

$\alpha(t)$ 为 t 时刻的学习率, 随着训练的进行逐渐减少.

$$\alpha(t) = \alpha_0 \exp\left(-\frac{t}{\tau}\right), \quad t = 0, 1, 2, \dots, \tau \quad (4)$$

其中, τ 为总的学习次数, α_0 为初始学习率.

由式(3)可知, SOM 算法对权值的调整, 仅考虑了学习率 α 及输入模式分量 x_j 与 $w_j(t)$ 之差, 即 $x_j - w_j(t)$, 而没有考虑在输入模式分量 x_j 与所有参与竞争的神经元权值 w 间实际存在的某种隐关系, 使系统缺乏整体的稳定性^[3]. 文献^[3]借助灰色系统理论刻画和描述 x_j 与 w 之间所存在的这种灰关系, 从而对 SOM 学习算法进行了改进, 构建了 GSOM. 然而, GSOM 忽略了输入模式与所有参与竞争的神经元权值间的整体特性.

整体特性是指输入模式和所有参与竞争的神经元各自作为一个整体间所存在的相似关系, 而不是输入模式分量与各神经元权值的关系. 本文采用了类似 P 范数的整体特性刻画. 由此, 本文建立了一个局部和总体相融合的灰关系, 并进一步使其一般化.

3 一般化的灰关系系数 (Generalized Grey Relation Coefficient, G^2RC)

对于一个系统, 研究一个给定参考模式与其它一组比较模式间的关系的方法, 称之为灰关系分析, 其中参考模式和比较模式都必须是归一化的模式. 下面以输入模式 \mathbf{X} 作为参考模式, 权值 $w_i (i=1, 2, \dots, M)$ 作为比较模式, 来分析 \mathbf{X} 和 w_i 间的灰关系. ξ_{ij} 称为输入模式分量 $x_j (j=1, 2, \dots, n)$ 和权值 w_{ij} 间的灰关系系数, 其定义为^[2]

$$\xi_{ij} = \frac{\Delta_{\min} + \lambda \Delta_{\max}}{\Delta_{ij} + \lambda \Delta_{\max}} \quad (5)$$

其中 $\lambda (0 \leq \lambda \leq 1)$ 是判别系数, 通常取 $\lambda = 0.5$, 并且

$$\Delta_{\min} = \min_{i=1, 2, \dots, M} \min_{j=1, 2, \dots, n} |x_j - w_{ij}| \quad (6)$$

$$\Delta_{\max} = \max_{i=1, 2, \dots, M} \max_{j=1, 2, \dots, n} |x_j - w_{ij}| \quad (7)$$

$$\Delta_{ij} = |x_j - w_{ij}| \quad (8)$$

显然, $0 < \xi_{ij} \leq 1$. 当 ξ_{ij} 越接近 1 时, Δ_{ij} 越接近

Δ_{\min} . 当 x_j 和 w_{ij} 间的相似度越高时, 即 Δ_{ij} 越接近 Δ_{\min} , w_{ij} 应更快地趋向于 x_j . 此时 ξ_{ij} 恰好体现了这种关系. 因此把 ξ_{ij} 引入到 SOM 的学习规则中将提高模型的性能, 从而有可能使算法具有更好的分类能力.

在上面的灰关系分析中, x_j 和 w_{ij} 之间的灰关系系数 ξ_{ij} 仅考虑了输入模式的单个分量 x_j 和权值 w_{ij} 之间的灰关系. 忽略了参考模式 \mathbf{X} 和比较模式 \mathbf{w}_i 各自作为一个整体所包含的一些重要信息. 因此, 对灰关系系数做如下修正:

$$\bar{\xi}_{ij} = \frac{\Delta_{\min} + \lambda \Delta_{\max}}{\Delta_{ij} + \lambda \Delta_{\max} + \eta \frac{1}{N} \left[\sum_{k=1}^N |x_k - w_{ik}|^p \right]^q}, \quad (9)$$

$0 \leq \lambda \leq 1; \eta \geq 0; p \geq 1, 0 < q \leq 1$

η 是一个非负实数, 它是 \mathbf{X} 和 \mathbf{w}_i 的整体特性的重要性度量, η 越大, 说明整体特性越重要, 若 $\eta=0$, 即忽略 \mathbf{X} 和 \mathbf{w}_i 的整体特性; 同时当 $\eta=0$ 时, $\bar{\xi}_{ij}$ 退化为 ξ_{ij} , 即 ξ_{ij} 是 $\bar{\xi}_{ij}$ 的一个特例, 因此改进后的 GRC 具有比 GRC 更好的描述特性. p, q 为正整数, 一般可取 $pq=1$, 即为 MINKOVSKI 距离, 度量 \mathbf{X} 和 \mathbf{w}_i 间的整体相近度. 显然, $0 < \bar{\xi}_{ij} \leq 1$, 且具有与式(4)定义的 ξ_{ij} 类似的特性.

为了使改进的灰关系系数更加一般化, 定义 $F(\bar{\xi}_{ij})$ 为灰关系系数 $\bar{\xi}_{ij}$ 的函数表示, 即一般化的灰关系系数 G^2RC (Generalized GRC). 对于 $F(\bar{\xi}_{ij})$ 函数可有多种形式的选择, 本文仅就如下 3 种形式进行讨论, 即指数形式、多项式形式、正切形式:

$$F(\bar{\xi}_{ij}) = \begin{cases} b \bar{\xi}_{ij}^{-1}, & b > 1 \\ \tanh(\beta \cdot \bar{\xi}_{ij} + \theta), & \beta, \theta > 0 \\ \bar{\xi}_{ij}^k, & k > 1 \end{cases} \quad (10)$$

$F(\bar{\xi}_{ij})$ 是对 $\bar{\xi}_{ij}$ 一般化的描述. 在第 5 节, 通过实验将证明选择 G^2RC 中不同函数, 可以解决不同的分类问题, 从而增强了 GRC 选择的灵活性.

4 G^2SOM 模型

在 SOM 算法中, 关于算法核心部分权值的调整由以下公式来决定:

$$\Delta w_{ij} = \alpha(t) (x_j - w_{ij}(t)), \quad i \in NE_{i^*}(t) \quad (11)$$

其中 $\alpha(t)$ 是学习率.

将一般化的灰关系函数引入 SOM 模型, 就构建了一个一般化的灰 SOM 模型 (G^2SOM).

由此, 式(11)中的权值调整量改为如下形式:

$$\Delta w_{ij} = \alpha(t) \times F(\bar{\xi}_{ij}) \times (x_j - w_{ij}(t)) \quad (12)$$

在上式中, $\eta=0$ 时, $\bar{\xi}_{ij}$ 退化为 ξ_{ij} , 此时若取 $F(\bar{\xi}_{ij}) = \bar{\xi}_{ij}^k$, G^2SOM 即为 GSOM. 由于在灰关系系数中加入了整体特性并使其一般化, 增强了 G^2SOM 模型对问题的描述能力.

如前所述, 在将灰关系系数引入 SOM 之前, 先对所有的输入向量和权矩阵作归一化处理. 样本归一化主要完成样本的预处理, 样本归一化公式为

$$x'_j = x_j / \sqrt{\sum_{i=1}^n x_i^2} \quad (13)$$

其中 $j=1, 2, \dots, n$, n 为输入模式 \mathbf{X} 的维数. 其本质是使归一化后的输入模式为单位长.

和 SOM 一样, G^2SOM 的学习过程同样包含了两个阶段即排序阶段和收敛阶段^[4]. 其中, 排序阶段是权矢量拓扑结构的调整阶段, 将初始学习率选为 0.9, 最终学习率为 0.05; 初始邻域为 8, 而最终邻域为 1; 而当算法进入收敛阶段, 算法需要提供输入空间的正确的统计特性, 从而进一步调整特征映射. 初始学习率为 0.05, 最终学习率为 0.01, 邻域范围始终为 0.

为了获得收敛, 排序阶段的学习的次数要选择 2000 次以上, 而进入收敛阶段, 学习次数可以在 1000 次以内, 以节省算法开销. 由于同 GSOM, SOM 同样的理论分析难度, 故与 GSOM 一样, 本文对 G^2SOM 采用 Benchmark 进行性能评估与比较, 实验包括 3 个数据集 Iris、同心圆和 Wine Recognition 数据集的分类试验.

同时, 在分类结果的评判中, 使用 Vote(投票表决)算法.

5 实验结果比较与评估

实验 1. 实验采用 Iris 数据集, Iris 数据集包括 3 个类, 第 1 类为 Setosa, 第 2 类为 Versicolor, 第 3 类为 Virginica; 每个类由 50 个 4 维模式组成; 并且第 2, 3 两类是线性不可分的. 使用 75 组数据训练, 75 组数据用于测试.

在实验中, 分别对 3 种灰关系函数进行了 10 次独立的无监督学习的分类实验. 其实验结果如表 1. 表 1 中列出了各个模型每次实验的错分模式数及平均错分数和平均错分率. GSOM 即采用文献[3]中的方法进行的实验. 同时, 为了保证实验的可行性、有效性和可比性对模式的一些参数作了如下的规定:

- (1) $m=5$, 二维神经元矩阵为 5×5 ;
- (2) 排序阶段:学习的最大次数为 2000;学习率的初始值为 0.9;学习率的最终值为 0.05;邻域的初始值为 8;邻域的最终值为 1;

- (3) 收敛阶段:学习的最大次数为 1000;学习率的初始值为 0.05;学习率的最终值为 0.01;邻域的初始值为 0;邻域的最终值为 0;

$$(4) b=2, k=4, \beta=1, \theta=\pi.$$

表 1 SOM, GSOM 及三种灰函数的 G^2 SOM 对 Iris 的聚类实验结果

模型	次数										平均错分数	平均错分率(%)
	1	2	3	4	5	6	7	8	9	10		
SOM	4	5	7	10	2	8	6	2	11	3	5.8	7.733
GSOM	6	11	2	9	3	3	2	2	5	6	4.9	6.533
指数	5	2	2	11	7	5	2	4	3	4	4.5	6.000
正切	2	3	3	7	0	5	5	1	3	6	3.5	4.667
多项式	6	4	4	4	5	2	4	3	5	0	3.7	4.933

而 b, k, β, θ 的选取是就问题而定的. 一般而言, 对于给定的大问题, 存在最优的参数 $b^*, k^*, \beta^*, \theta^*$, 使该问题有最优解.

由表 1, 正切函数 $F(\xi_{ij}) = \tanh(\beta \cdot \xi_{ij} + \theta)$ 对数据的分类问题效果最好, 错分率比 GSOM 下降了 1.866%, 但由于参数选择的多样性, 其它的灰关系函数形式对不同的问题通过参数的调整, 可能有更好的效果. 同时, 二维神经元矩阵设为 10×10 时, 识别率更低, 为 2.04%. 由以上结果可知参数选择是 G^2 SOM 模型的关键.

实验 2. 人工产生两组 2 维同心圆数据 A, 加部分随机噪声. 样本数为 100, 两类各 50, 如图 2 所示.

由图 2 可知十字点和圆点是非线性可分的. 为了提高用 G^2 SOM 对同心圆数据的分类效果, 在对数据进行分类前, 首先构造一个核映射, 该映射从低维空间映射到高维空间, 即由 (x_1, x_2) 映射到 $(y_1, y_2, y_3, y_4, y_5)$, 采用多项式形式的内积函数:

$$K(x, x_i) = \frac{1}{4} [(2(x \cdot x_i) + 1)^q - 1], \quad q = 2 \quad (14)$$

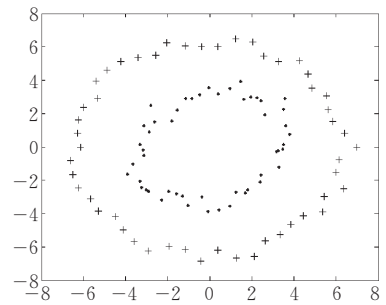


图 2 同心圆数据集

记映射 $\phi: R^2 \mapsto R^5$,

$$(x_1, x_2)^T \xrightarrow{\phi} (y_1, y_2, y_3, y_4, y_5)^T = (x_1, x_2, \sqrt{2} \cdot x_1 \cdot x_2, x_1^2, x_2^2)^T \quad (15)$$

数据 A 经过多项式坐标变换核函数的非线性映射函数映射后, 得到 5 维的数据 B, 每次随机从每一类样本中抽取 25 个用于训练, 剩下 25 个用于测试, 即 50 组数据训练, 50 组数据用于测试. 同样, 分别对上述三种灰关系函数分别进行了 10 次分类实验. 其实验结果如表 2 所示.

表 2 SOM, GSOM 及三种灰函数的 G^2 SOM 对同心圆聚类的实验结果

模型	次数										平均错分数	平均错分率(%)
	1	2	3	4	5	6	7	8	9	10		
SOM	4	4	2	4	0	0	0	0	1	1	1.6	3.2
GSOM	5	0	0	2	0	0	2	4	2	0	1.5	3.0
指数	0	2	0	1	0	1	0	1	0	2	0.7	1.4
正切	1	0	0	3	0	3	2	1	1	2	1.3	2.6
多项式	2	1	2	0	0	2	0	5	2	4	1.8	3.6
多项式($k=3$)	5	3	1	1	0	0	2	0	0	1	1.3	2.6

从表 2 可以看出, G^2 SOM 对同心圆数据的聚类效果比传统的 SOM 要好, 三种函数形式中, 指数和正切形式的灰函数效果较好, 错分率分别下降了 1.8% 和 0.6%, 而当 $k=4$ 时, 多项式形式的效果较差, 当 $k=3$ 时, 错误率为 2.6%, 这说明和实验 1 一样, 系统的识别率依赖于参数的选择.

实验 3. Wine Recognition 数据集^[3]对三种意大利葡萄酒的化学分析结果. 数据集有 178 个样本, 每个样本 13 维特性. G^2 SOM 初始参数设置与 Iris 数据集相同. 采用 90 个样本训练, 88 个样本测试, 实验结果如表 3 所示.

表 3 GSOM 和三种灰函数的 G²SOM 对 Wine Recognition 的实验结果

模型	次数										平均错分数	平均错分率(%)
	1	2	3	4	5	6	7	8	9	10		
GSOM	16	15	18	20	22	23	24	17	20	21	19.6	22.2
指数	17	22	21	12	17	18	20	20	19	19	18.5	21.0
正切	20	16	19	18	14	17	20	17	22	15	17.8	20.2
多项式	19	20	17	19	17	24	21	19	15	19	19.0	21.6

从表 3 可以看出, G²SOM 对 Wine Recognition 数据的聚类效果比 GSOM 要好, 三种函数形式对 GSOM 的错分率分别下降了 1.2%, 2.0% 和 0.8%。

6 结 论

本文利用灰关系系数及其一般化形式得出了基于灰关系函数的 SOM 改进模型 G²SOM. 目的在于将神经元间的内在关系通过灰关系函数来体现, 并融入到 SOM 的学习规则中, 进一步提高算法的性能. 实验验证了 G²SOM 的有效性, 它降低了系统相应的错分率, 提高了识别的效果. 进一步的工作是将 G²SOM 引入入侵检测 (IDS) 中的异常检测部分, 提高现有基于规则的 IDS 的自学习和检测能力.

参 考 文 献

- Huang D. S. . Theory of Neural Network and Pattern Recognition System. Beijing: Publishing House of Electronics Industry, 1996 (in Chinese)
(黄德双. 神经网络模式识别系统理论. 北京: 电子工业出版社, 1996)
- Deng J. L. . Control problems of grey systems. Systems Con-

- trol Letters, 1982, 1(5): 288~294
- Hu Yi-Chung, Chen Ruey-Shun, Hsu Yen-Tseng, Tzeng Gwo-Hshung. Grey self-organizing feature maps. Neurocomputing, 2002, 48(4): 863~877
- Haykin Simon. Neural Networks: A Comprehensive Foundation. Beijing: Tsinghua University Press, 2001
- Carpenter G. A. , Grossberg S. . A massively parallel architecture for a self-organizing neural pattern recognition machine. Computer Vision, Graphics, Image Process, 1987, 37(1): 54~115
- Carpenter G. A. , Grossberg S. , Markuzon N. , Reynolds J. H. , Roson D. B. . Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. IEEE Transactions on Neural Networks, 1992, 3(5): 698~713
- Cheng C. S. , Hsu Y. T. , Wu C. C. . Grey neural networks. IEICE Transactions on Fundamentals Electronics, Communication Computing Science, 1998, E81-A(11): 2433~2442
- Weijters A. . The BP-SOM architecture and learning rule. Neural Processing Letters, 1995, 2(6): 13~16
- Su M. C. , Chang H. T. . A new model of self-organizing neural networks and its application in data projection. IEEE Transactions on Neural Networks, 2001, 12(1): 153~158
- Su M. C. , Chang H. T. . Fast self-organizing feature map algorithm. IEEE Transactions on Neural Networks, 2000, 11(5): 721~733



PAN Zhi-Song, born in 1973, Ph. D. candidate. His research interests include network security, intrusion detection system and neural network.

CHEN Song-Can, born in 1962, professor and Ph. D. supervisor. His main research interests include neural network and pattern recognition.

ZHANG Dao-Qiang, born in 1978, Ph. D. candidate. His main research interests include neural network and pattern recognition.

Background

The project name is Improved SOM and their Performance Evaluations. The project is partial supported by National Science Foundations of China. The aim of the project is how to improve the classification performance of SOM using

kernel means and grey theory. In the paper, Generalized Grey Self-organizing Maps is designed and the experiments on the benchmark datasets confirm their validities and feasibilities.