

# 基于模糊神经网络的数据挖掘算法

赵高峰<sup>1</sup>, 毕笃彦<sup>1</sup>, 孙卫<sup>2</sup>

(1. 空军工程大学 工程学院, 陕西 西安 710038; 2. 西安交通大学 系统工程研究所, 陕西 西安 710049)

**摘要:**提出一种基于模糊神经网络进行数据挖掘的新方法。构成模糊神经网络的模糊化层采用高斯函数计算5个模糊隶属度,高斯函数需要的均值、方差以及隶属度的中心值都可通过预先计算采集到的数据得到。模糊推理层采用取小取大运算代替常用的积和运算,加快了网络的推理速度。在模糊神经网络训练阶段,首先利用重心法对模糊化层输出进行反模糊化,再采用BP思想,利用梯度法求误差值并进行反传调整隶属度函数的参数值。为提高网络推理精度和速度,通过设立相应的规则对网络进行裁剪,剪掉多余的节点和权值,最后依据一定的思想对产生的模糊规则进行简化和提取。以工业锻造中的智能温度控制系统的控制数据为例进行仿真,结果表明,该网络具有较高的精度和较好的鲁棒性。

**关键词:**数据挖掘;模糊神经网络;网络裁剪;规则提取

**中图分类号:** TP31 **文献标识码:** A **文章编号:** 1009-3516(2008)03-0063-04

数据挖掘是指从大量数据中提取出未知的、有价值的模式或规律。现已经应用在工业控制、商业管理、科学研究等领域。数据挖掘的算法有许多种,主要包括:模糊集,粗糙集,神经网络等<sup>[1]</sup>。现实中巨量数据集中数据关系复杂,非线性程度高,噪声数据普遍存在,使得一些算法的适用受到很大的限制。

人工神经网络能够通过训练获得用数据表达的知识<sup>[2-3]</sup>,除了可以记忆已知的知识外,神经网络还具有较强的概括能力和联想记忆能力,不错的容噪能力和分类精度、鲁棒性高的优点使它在解决数据挖掘中的问题时有很大的优势。但它也有一些不足,如“黑箱”性,对于高维或者超大规模问题,它的学习速度缓慢,规则生成比较困难<sup>[4]</sup>。

模糊逻辑(Fuzzy Logic)是以 Zadeh L A 创立的模糊集合论为理论基础,针对人类思维中的模糊性特点。利用领域专家的先验知识,进行近似推理。它能对复杂事务进行模糊识别,模糊推理,但缺乏自学习自适应的能力。

将模糊逻辑引入人工神经网络实现它们的有机结合,形成模糊神经网络(Fuzzy Neural Network, FNN),使系统不仅具有模糊逻辑的不确定信息处理能力,又具有神经网络良好的自学习性。和一般的神经网络相比,它具有更快的学习速度和更小的规模。

## 1 模糊神经网络实现

### 1.1 模糊神经网络结构

模糊神经网络就是用模糊化模块对数据进行处理,然后将获得的知识传递到神经网络中进行网络训练,最后将结构进行清晰化处理并输出<sup>[5]</sup>。模糊神经网络系统的结构如图1所示。

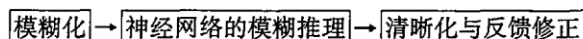


图1 模糊神经网络系统结构

Fig. 1 Fuzzy neural network system structure

收稿日期:2007-05-29

基金项目:国家“863”计划资助项目(2006AA701121)

作者简介:赵高峰(1981-),男,陕西扶风人,硕士生,主要从事数据挖掘研究;E-mail:gaofenggj@163.com

毕笃彦(1963-),男,陕西扶风人,教授,博士,主要从事图像处理和模式识别研究。

## 1.2 模糊神经网络构造

依照图1构建了一种模糊神经网络。该网络共有8层。

第1层为输入层,包括固定值-1输入和样本变量值 $x$ 的输入;第2层为偏移层,计算 $x_i - \omega_{ij}^{\mu}$ ;第3层为模糊输入层(或者称为高斯函数层)。该层采用高斯函数作为传递函数主要用于实现模糊隶属度函数,计算出5个分属于“差,较差,可,较好,好”语言变量的隶属度;第4层为取小运算层,完成输入值的取小运算;第5层为取大运算层;第6层为模糊输出层,将 $P$ 结点的输出和此层的其它结点输出相乘输出;第7层为清晰化层,主要完成对映射网络的输出结果进行清晰化处理;第8层为权值调整层。

## 1.3 隶属度函数实现

这任务由前3层来实现对模糊信息的预处理,其主要功能是计算各输入分量属于各模糊集合的隶属度。

设输入向量为 $X = [x_1, x_2, \dots, x_n]^T$ ,则模糊化层各节点的输入输出如下:

$$\text{第1层: } I_i^1 = x_i, O_i^1 = I_i^1 \quad \text{第2层: } I_{ij}^2 = O_i^1 - \omega_{ij}^{\mu}, O_{ij}^2 = I_{ij}^2$$

$$\text{第3层: } I_{ij}^3 = O_{ij}^2 \cdot \omega_{ij}^{\sigma} \quad O_{ij}^3 = f(I_{ij}^3) = \exp\{-((I_i^1 - \omega_{ij}^{\mu}) / (1/\omega_{ij}^{\sigma}))^2\}$$

其中, $i=1, 2, \dots, n, j=1, 2, 3, 4, 5$ 。 $I_k^i, O_k^i$ 分别代表第 $k(k=1, 2, 3)$ 层第 $i$ 个节点的输入输出。 $I_{ij}^k, O_{ij}^k$ 分别代表 $k$ 层中由第 $i$ 个输入节点分出的第 $j$ 个节点的输入和输出。 $\omega_{ij}^{\mu}, \omega_{ij}^{\sigma}$ 为可调函数, $\omega_{ij}^{\mu}$ 为高斯函数的中心值; $\omega_{ij}^{\sigma}$ 为函数的分布参数,又称为尺度因子。

设样本的均值和方差分别为 $\mu$ 和 $\sigma$ ,由统计知识知:

$$\mu = 1/n \left( \sum_{i=1}^n x_i \right) \quad \sigma = \sqrt{1/n \left( \sum_{i=1}^n (x_i - a)^2 \right)}$$

上式中, $n$ 为样本数。在对数据进行模糊分类时,取“差”、“较差”、“可”、“较好”的隶属函数的中心分别取为 $(\mu - 2\sigma)$ 、 $(\mu - \sigma)$ 、 $\mu$ 、 $(\mu + \sigma)$ 、 $(\mu + 2\sigma)$ 。

## 1.4 模糊神经网络训练

当系统的推理精度不够高时,需要调整隶属函数的参数<sup>[6]</sup>。采用BP网络的反向传播学习思想进行训练。对于第 $i$ 组采样数据 $(x_1, x_2, \dots, x_n; y_1, y_2, \dots, y_m)^T$ 。训练算法如下:

第1步:提取模糊规则。将样本输入到网络后,在第3层中计算“差,较差,可,较好,好”的语言值的隶属度。先将这5个语言值标记为1,2,3,4,5,对于每一个输入元素,取出其中隶属度最大的记录下来,在相应语言值下面标上1,其余的记为0。最后,按照对应元素的顺序将规则表示出来,当有重复规则时,用CF值表示,有一个重复规则,CF自加0.1。

第2步:清晰化。第7层中按重心法<sup>[7]</sup>计算反模糊化的输出 $y_i^* = (\sum_{k=1}^5 O_k^6 \cdot \omega_p^k) / (\sum_{k=1}^5 O_k^6)$ 其中, $\omega_p^k$ 为第6层中第 $k$ 个节点和第7层节点之间的连接权值,为清晰化参数, $i$ 为第 $i$ 组样本。

第3步:误差反传。在训练网络时采用了梯度法,用交叉熵函数作为误差函数,计算推理误差,并修正参数。其形式为: $F(\cdot) = - \sum_{i=1}^n \sum_{k=1}^m y_r^d \ln y_r^* + (1 - y_r^d) \ln(1 - y_r^*)$ ,式中, $y_r^d$ 为第 $r$ 组样本在第7层期望输出值, $y_r^*$ 为第 $r$ 组样本在第7层的实际输出值。

## 1.5 网络裁剪

网络训练成功后,一般会存在一些冗余权,可依据一定的规则进行裁剪<sup>[8]</sup>,使网络权和节点的数目最小。裁剪算法如下:

- 1) 设 $\eta_1$ 为网络的误差极限, $\eta_2$ 为权值删除的阈值,且 $\eta_1 + \eta_2 < 0.5$ ,训练网络到预定精度;
- 2) 对于网络中的权 $\omega$ ,若 $\max_k |\omega_k^p \omega| \geq 4\eta_2$ ,则删除此权值 $\omega$ ;对于网络中的权 $\omega_k^p$ ,若 $|\omega_k^p| \geq 4\eta_2$ ,则删除此权值 $\omega_k^p$ ;
- 3) 若无法满足2)的权值,计算 $\eta = \max_k |\omega_k^p \omega|$ ,删除最小的 $\eta$ ;
- 4) 再次训练网络,若网络的精度小于预定的精度,返回到3)中取舍网络权值,否则,停止,用此时的权值。

网络收敛是以精度达到给定的误差极限为条件。这种算法对有许多冗余权值的全连结网络尤其适用<sup>[9]</sup>。

### 1.6 模糊规则简化

网络训练成功后,会产生很多规则。这些规则中,有我们感兴趣的,也有不感兴趣的,需要对规则进行简化<sup>[10]</sup>。简化的思想是:①将网络权值为0的规则或者很接近0的规则视为无效规则;②在具有同一前提的所有规则中,将权值最大的规则保留,其它的规则去掉。

## 2 实例仿真

以工业锻造淬火工艺中智能温度控制系统用于物料温控的控制数据为例,通过上面提出的方法形成温度控制模糊规则进行控制仿真,以验证本方法的有效性。

通过对工业历史资料分析,提取了两个输入因子,1个输出因子的网络。以温度误差  $T$  和温度误差变化  $T_k$  为输入因子,以控制量  $Y$  为输出因子。学习效率  $\beta$  取值为 0.45。

对样本数据进行分析后,以最佳温度为中心,以其它温度与它差值的均分值作为  $T$  和  $T_k$  的论域: $T = T_k = \{-6, -4, -2, 0, 2, 4, 6\}$ ,  $Y$  的论域为: $Y = \{-6, -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, 6\}$ 。选取  $T$  和  $T_k$  的模糊集合元素为:负大,负中,接近于零,正中,正大。 $Y$  的模糊分为7个部分:负大,负中,负小,零,正小,正中,正大。

仿真通过 Matlab 的图形界面进行操作,仿真中,首先给 FNN 网络的各参数和权值赋以初值,使用 252 条数据作为网络的学习样本。隐层中,以具有最大隶属度为 1 的样本值作为高斯函数的中心值,宽度取  $[-6, 6]$  中的整数。经过 842 次训练后,网络收敛。裁剪后,网络裁剪了权  $\omega_2^j$ 。再次训练网络,收敛后,  $T, T_k$  的高斯函数中心值由最初的  $-3, -1, 0, 1, 3$  分别向隶属度函数中心  $-4, -2, 0, 2, 4$  偏移,隶属函数比训练前更平坦一点训练前后的隶属度函数如图 2 所示。

根据训练样本和收敛的网络,对测试样本进行规则提取以检查网络控制的有效性,经过网络裁剪,规则提取,产生出部分规则。结果显示,控制结果的最大误差为 0.118 6,平均误差小于 0.093 7,证明了算法的有效性。

用传统的 BP 网络对上例进行仿真,与本文方法进行对比。对比结果如图 3 所示,文章提出的方法的控制性能优于传统的 BP 网络。由此可见,将模糊逻辑和神经网络结合起来,可提高系统数据挖掘的效率及提取规则的可靠性,提高系统的控制精度和鲁棒性。

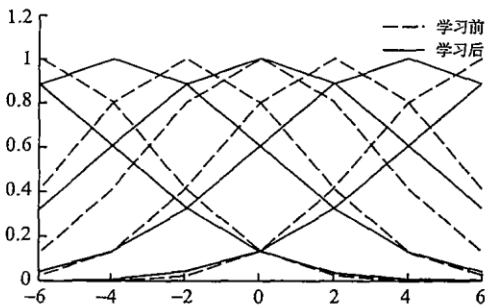


图2 网络训练完成后隶属函数曲线图  
Fig. 2 Membership chart after the net train successfully

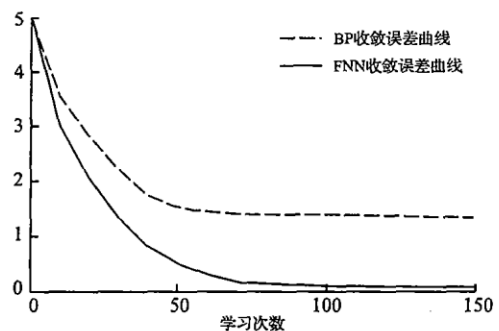


图3 FNN与BP收敛精度对比  
Fig. 3 Precision comparison between FNN and BP

## 3 结束语

将模糊系统和神经网络结合起来,构造一种模糊神经网络,实现模糊规则的自动生成,提高了模糊神经网络的自学习性能和鲁棒性。利用 BP 思想实现网络权值的调整,加快了网络的收敛速度。对规则的简化,取出冗余规则,从而挖掘尽可能接近实际的完全规则,使得所挖掘的规则简单、可靠。仿真结果表明文章提出的方法具有控制精度高,鲁棒性好的优点。

## 参考文献:

- [ 1 ] 陈 莉. 数据库中的知识发现[J]. 西北大学学报:自然科学版,1999,29(1):114-119.  
CHEN Li. Knowledge Discovery in Database[J]. Journal of Northwest University:Natural Science Edition,1999,29(1):114-119. (in Chinese)
- [ 2 ] 曲东才. 增强神经网络辨识模型泛化能力的研究[J]. 海军航空工程学院学报,2007,22(1):109-113.  
QU Dongcai. Design and Simulation of the Improved Performance Index Function with Enhanced Generalization Ability for the ANN's Identification Model[J]. Journal of Naval Aeronautical Engineering Institute,2007,22(1):109-113. (in Chinese)
- [ 3 ] 韩 虎,张莉莉. 神经网络方法在数据挖掘中的应用[J]. 系统工程学报,2004,23(4):88-90.  
HAN Hu,ZHANG Lili. Application of Neural Network in Data Mining[J]. Journal of Systems Engineering,2004,23(4):88-90. (in Chinese)
- [ 4 ] 王海龙. 协同神经网络在图像识别中的应用研究[D]. 上海:上海交通大学,2000.  
WANG Hailong. Analysis of the Recognition Performances of Synergetic Neural Network[D]. Shanghai:Shanghai Jiaotong University,2000. (in Chinese)
- [ 5 ] 韩 敏,孙燕楠. 多输入模糊神经网络及其应用[J]. 系统工程与电子技术,2003,25(10):1249-1252.  
HAN Min,SUN Yannan. Multi-Input Fuzzy Neural Network and Its Application[J]. Systems Engineering and Electronics,2003,25(10):1249-1252. (in Chinese)
- [ 6 ] LEEA H C. Neural Network Architecture for Pattern Recognition and Data Mining[J]. IEEE Trans on Systems,Man and Cybernetics,2003,11:834-838.
- [ 7 ] Hong Jun Lu,Rudy Setiono,Huan Liu. Effective Data Mining Using Neural Networks[J]. IEEE Trans on Systems,Man and Cybernetics,2001,24:1279-1284.
- [ 8 ] Bezdek J C,Pal S K. Fuzzy Models for Pattern Recognition[M]. [S. l.]:IEEE CS Press,1999.
- [ 9 ] Bezdek J C. Pattern Recognition with Fuzzy Objective Function Algorithms[M]. New York:Plenum Press,1999.
- [ 10 ] Krishnapanam R,Killer J M. A Possibilistic Approach to Clustering[J]. IEEE Trans on Fuzzy System,2000,1(2):98-105.

(编辑:田新华,徐楠楠)

## A Data Mining Algorithm Based On Fuzzy Neural Net

ZHAO Gao-feng<sup>1</sup>, BI Du-yan<sup>1</sup>, SUN Wei<sup>2</sup>

(1. Engineering Institute, Air Force Engineering University, Xi'an 710038, China; 2. System Engineering Institute, Xi'an Jiaotong University, Xi'an 710049, China)

**Abstract:** An algorithm of Data Mining based on fuzzy neural net is proposed in the article. The algorithm creates a new fuzzy neural net, in which Gaussian-Function is used to calculate five fuzzy memberships in fuzzy layer. The average and variance required by Gaussian-Function can be calculated in advance through the data to be collected. Using of fuzzy inference from a Max-min operation to replace commonly multiply-add operation accelerates the speed of the network. In training stage, first the centre-of-gravity method is used to resist the fuzzy of the output of the fuzzy layer, and then BP idea is adopted to calculate the error and adjust the membership function parameters. To improve the accuracy and speed of fuzzy neural network, the net, using appropriate rules, crops redundant nodes and weights to extract and simplify the rules. A simulation is performed by using the control data from the intelligent temperature control system in industrial forging, the result shows that the algorithm has higher precision and stronger robustness.

**Key Words:** data mining; fuzzy neural net; net crop; rules extraction