神经网络 模糊联想记忆 学习算法 (17)

计算机科学1994 Vol. 21 №. 3

67~68

多模式对模糊联想记忆学习算法的优化

TP18

何奉道

（西南交通大学计算机科学与工程系　成都 610031）

摘　要　this paper proposed a differential rule of fuzzy Max-Min composition operation. With a cost function reflecting the features of neural network, combining the gradient descent search technique with fuzzy Hebb rule, an optimization algorithm for associated to store multiple pattern pairs in a single FAM has been established.

关键词　Neural network, Fuzzy associative memory, Fuzzy Hebb rule, Optimization.

一、引言

人脑神经系统信息活动的一个重要特征是能够接收和处理模糊的、连续随机的信息，并在输出时不追求绝对精确解而只要求能找到问题的满意解。模糊联想记忆神经网络是 Bart Kosko 于1987年提出的采用模糊赫布型学习规则的一种单状态异联想记忆神经网络，在模糊控制、模式识别、专家系统等领域曾引起人们的关注。由于该网络的子集联想特性和不能有效地联想存储多个训练模式对而影响了它的应用[1~3]。本文提出模糊合成运算的一种微分法则，用一代价函数以反映网络性能，将梯度下降搜索技术与模糊赫布型学习规则相结合，建立了在单个模糊联想记忆神经网络中联想存储多个模式对的一种优化学习算法。理论分析和实例计算均证明该算法优于 Kosko 的学习规则。

二、模糊赫布型学习规则

模糊联想记忆神经网络是一种能联想存储任意模糊值(0～1)模式对(A,B)，$A=(a_1,a_2,\cdots,a_n)$，$B=(b_1,b_2,\cdots,b_m)$的两层神经网络，输入层和输出层神经元广泛连接，输入层的 n 个单元对应于 A 的 n 个分量，输出层的 m 个单元对应于 B 的 m 个分量，神经元 a_i 和 b_j 间的连接权为 w_{ij}。

Kosko 采用模糊赫布型学习规则定义网络关于模式对(A,B)的连接权矩阵 $W=[w_{ij}]$，其中

$$w_{ij}=a_i\wedge b_j,\ i=1,2,\cdots,n;\ j=1,2,\cdots,m \tag{1}$$

它反映出神经元 a_i 和 b_j 之间的联系随它们同时兴奋而增强。

网络的联想由方程

$$B'=A\circ W \tag{2}$$

完成。式中 ∘ 为模糊极大-极小合成运算。即

$$b'_j=\bigvee_{i=1}^{n}(a_i\wedge w_{ij}),\ j=1,2,\cdots,m \tag{3}$$

∧运算反映了输入刺激量 a_i 与权 w_{ij} 之间的限制性修正作用，∨运算则反映了神经元之间的主元素突出型竞争方式，合成则是一种主因素决定型合成关系。

当向网络输入任意模糊值模式 A'，设联想输出模式为 B'，则

$$b'_j\leqslant\bigvee_{i=1}^{n}w_{ij}\leqslant\bigvee_{i=1}^{n}b_j=b_j \tag{4}$$

即是说，对于任意的模糊值输入模式，网络的联想输出总是已存储模糊集的子集，这是采用模糊赫布型学习规则的模糊联想记忆神经网络的主要缺点。

为了在单个网络中联想存储多个模式对(A^k,B^k)，$k=1,2,\cdots,p$，文[2]提出一种学习规则：

$$w_{ij}=\bigvee_{k=1}^{p}w_{ij}^{k}=\bigvee_{k=1}^{p}(a_i^k\wedge b_j^k);$$
$$i=1,2,\cdots,n;\ j=1,2,\cdots,m \tag{5}$$

网络的回想或联想由下面方程完成：

$$B^k=A^k\circ W,\ k=1,2,\cdots,p \tag{6}$$

显然，多模式对模糊联想记忆学习规则可归结为寻求使(6)式成立的模糊关系矩阵。然而在大多数情况下，这样的矩阵是不存在的。即使存在，由(5)式所获得的权矩阵也不一定是解，另一方面，由训练模式对集合通过一次学习便形成网络的连接权矩阵。

这样，训练模式对的选择对权矩阵的形成关系极大，而要求给出理想的训练模式对往往是困难的。为使网络具有更好的回想和联想特性，一个有效的方法是对由(5)式形成的权矩阵进行自适应优化。

三、优化算法

我们首先建立网络性能的评价标准。设 $\bar{b}_j^k$ 为 B^k 的第 j 个分量，$\bar{b}_j^k$ 为网络对应单元的实际输出，第 k 个模式对的回想误差函数为：

$$E_k=\frac{1}{2}\sum_{j=1}^{m}(\bar{b}_j^k-b_j^k)^2,\ k=1,2,\cdots,p \tag{7}$$

即

$$E_k=\frac{1}{2}\sum_{j=1}^{m}[\bigvee_{i=1}^{n}(a_i^k\wedge w_{i,j})-b_j^k]^2,\ k=1,2,\cdots,p \tag{8}$$

整个训练模式对集合的全局代价函数为

$$E=\sum_{k=1}^{p}E_k \tag{9}$$

调整网络连接权使 E 最小，这可由梯度下降规则来实现：

$$\Delta W=-\alpha(t)\left[\frac{\partial E}{\partial w_{i,j}}\right] \tag{10}$$

其中，$\alpha(t)$为学习率函数，$0<\alpha(t)\leqslant 1$。

(10)式中梯度的计算含有模糊合成运算的微分。为此，我们定义下面几种模糊算子的微分规则：

(1)设 a、b 为两模糊变量，$\wedge$、$\vee$ 分别为模糊极小、极大算子，则

$$\frac{\partial}{\partial a}(a\wedge b)=\begin{cases}1 & 若\ a\leqslant b,\\ 0 & 否则。\end{cases} \tag{11}$$

$$\frac{\partial}{\partial a}(a\vee b)=\begin{cases}1 & 若\ a\geqslant b,\\ 0 & 否则。\end{cases} \tag{12}$$

(2)设 X、Y 为两 n 维模糊向量，$\circ$ 为模糊 $\vee-\wedge$ 合成算子，则

$$\frac{\partial}{\partial x_i}(X\circ Y^T)=\frac{\partial}{\partial x_i}[\bigvee_{j=1}^{n}(x_j\wedge y_j)]$$

$$=\begin{cases}1 & 若\ x_i=(x_i\wedge y_i)\geqslant \bigvee_{j=1,j\neq i}^{n}(x_j\wedge y_j),\\ 0 & 否则。\end{cases} \tag{13}$$

这样

$$\frac{\partial E}{\partial w_{i,j}}=\sum_{k=1}^{p}\frac{\partial E_k}{\partial w_{i,j}}$$

$$=\sum_{k=1}^{p}\{[\bigvee_{r=1}^{n}(a_r^k\wedge w_{r,j})-b_j^k]\cdot\delta_{i,j}\}$$

$$\delta_{i,j}=\frac{\partial}{\partial w_{i,j}}[\bigvee_{r=1}^{n}(a_r^k\wedge w_{r,j})-b_j^k]$$

$$=\begin{cases}1 & 若\ w_{i,j}=(a_i^k\wedge W_{i,j})\geqslant \bigvee_{r=1,r\neq i}^{n}(a_r^k\wedge w_{r,j}),\\ 0 & 否则。\end{cases} \tag{14}$$

由(10)式，网络连接权矩阵的变化应正比于训练模式对集合上各模式对对应的负梯度之和。为简化计算，可每给网络提供一个模式对，即计算误差函数 E_k 的梯度并调整连接权矩阵，这样虽偏离了全局误差函数 E 上真正的梯度下降，但当 $\alpha(t)$ 较小时，这种偏离是可以忽略的。若以模糊赫布型学习规则获得的权矩阵为迭代初值，网络将收敛于全局误差函数曲面的极小值点。

综上所述，多模式对模糊联想记忆优化学习算法可描述如下：

(1)初始化：由(5)式设置初始权矩阵 $W^{(0)}$；给定精度 ε 和迭代次数上限 μ。

(2)对每一模式对 (A^k,B^k)，$k=1,2,\cdots,p$，调整网络连接权矩阵：

$$W^{(t+1)}=W^{(t)}-\alpha(t)\left[\frac{\partial E_k}{\partial w_{i,j}^{(t)}}\right]$$

$$\frac{\partial E_k}{\partial w_{i,j}^{(t)}}=[\bigvee_{r=1}^{n}(a_r^k\wedge w_{r,j}^{(t)})-b_j^k]\cdot\delta_{i,j}^{(t)}$$

$$\delta_{i,j}^{(t)}=\frac{\partial}{\partial w_{i,j}^{(t)}}[\bigvee_{r=1}^{n}(a_r^k\wedge W_{r,j}^{(t)}-b_j^k]$$

$$=\begin{cases}1 & 若\ w_{i,j}^{(t)}=(a_i^k\wedge w_{i,j}^{(t)})\geqslant \bigvee_{r=1,r\neq i}^{n}(a_r^k\wedge w_{r,j}^{(t)}),\\ 0 & 否则。\end{cases} \tag{14}$$

(3)计算 E。

(4)若 $E<\varepsilon$ 或 $t>\mu$ 则结束，否则 $t=t+1$，返回(2)。

四、结束语

众所周知，神经网络能否较准确地记忆学习样本，是决定其性能优劣的一个关键因素。某一类问题确定之后，通过特征抽取可得到该类问题的典型样本集合。在模糊控制、模糊推理、模式识别、专家系统等领域，典型样本均为或可转化为模糊值模式对。模糊联想记忆神经网络是将模糊集合理论和神经网络分布式存储和大规模并行处理机制相结合的一种信息处理系统，因而在上述领域有着良好的应用前景。针对该网络不能有效地联想记忆训练样本，我们提出一种将梯度下降技术和模糊赫布型学习规则相结合的优化学习算法，使外部环境信息通过内部自组织更为有效地适应环境。(参考文献共5篇略)