人工 神经网络 容限空间理论 扩展

计算机科学 1994 Vol. 21 №. 5

27-30

容限空间理论的扩展及其在人工神经网络中的应用

毛国君 (大连理工大学工程力学研究所)
钱 弢 吕建刚 (徐州师范学院数学系)

TP18

摘 要 TST (Tolerance Spaces Theory) suggested by Zmoeeman has defined some concepts such as Tolerance and Tolerance Space. Based on known knowledge on TST, this paper mainly discusses the following problems: 1) the further study of TST, specially in the definition of Generations; 2) TST is applied to some problems of Artificia Neural Networks; 3) an algorithm which can be used for machine learning is given.

关键词 Tolerance spaces theory, Artificia neural network, Machine learning, AI.

一、基本概念

我们知道，在一般的代数空间中，集合 X 和 Y 的二元关系定义为 X×Y 的子集。Zmoeeman 在研究人的视觉结构中，从一般拓扑空间理论出发建立了容限空间理论。在容限空间理论中，定义了容限、容限空间等概念。这些概念的提出对于我们正确认识人工神经网络功能有着重要的指导意义。

定义 1 集合 X 上的一个容限 ξ 是 X 上的一个二元关系，即 $\xi\subseteq X\times X$，且满足

(i)反身性：$\forall x\in X, (x,x)\in\xi$；

(ii)对称性：$\forall x,y\in X$，若 $(x,y)\in\xi$，则 $(y,x)\in\xi$。

定义 2 一个容限空间是一个具有容限的集合。X 上具有容限 ξ 的容限空间记为 (X,ξ)。

定义 3 对于容限空间 (X,ξ)，容限 ξ 的 2 阶容限定义为：

$\xi^2=\{(x,y)|\exists x_i\in X$，使得 $(x,x_i)\in\xi$ 且 $(x_i,y)\in\xi\}$；

一般地，当 $n>2$ 时，容限 ξ 的 n 阶容限递归定义为

$\xi^n=\{(x,y)|\exists x_i\in X$，使得 $(x,x_i)\in\xi,(x_i,y)\in\xi^{n-1}\}$.

定义 4 X 中关于它的某个元素 X 的连通分支 C(x) 是指

$$C(x)=\{y|\exists \text{ 正整数 } n, \text{使得} (x,y)\in\xi^n\}.$$

定义 5 若对 $\forall x\in X$，有 $C(x)=X$，则称 (X,ξ) 是连通的。

容限空间的连通性可用以指导人工神经网络的功能实现，为了说明这个问题，我们看一下分类功能。设被观察客体 X 包含 M 类可区分的子客体，记为 $C_i(i=1,2,\cdots,M)$，一个分类问题就是依据某个输入

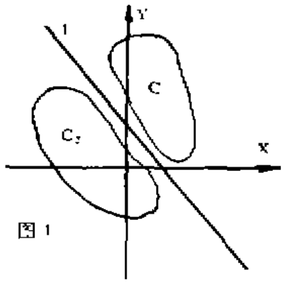

图 1

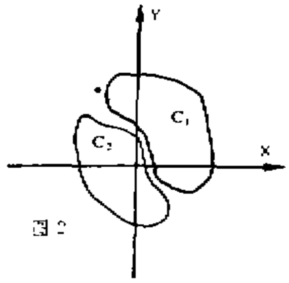

图 2

x 找出 x 所属的类 C_i。如果从容限空间的连通性角度来看，人工神经网络的这种分类功能就是寻找 x 所在的连通分支 C(x)，使之与某个 C_i 相等(或者是相差无几)。例如，图 1 所示的简单情况，被观察客体是二维图形，只有两类 C_1 和 C_2，而且可用一条直线 $l: ax+by-c=0$ 隔开。如果对点 $p_1(x_1,y_1)$ 和 $p_2(x_2,y_2)$，定义 $(p_1,p_2)\in\xi$ 当且仅当 ax_1+by_1-c 和 ax_2+by_2-c 同号，那么对任意输入点 (x,y)，若 $ax+by-c>0$，则 (x,y) 所在的连通分量是 C_1，否则为 C_2。对于类似于图 2 所示的 C_1 和 C_2 不能用直线线性分隔的情况，寻找输入点所在的类要麻烦些。一般要采用最小二乘学习算法来进行，即给定预期输出量 $d(k)$，通过寻求它与实际输出量 $y(k)$ 的统计均方差 $E[(d(k)-y(k))^2]$ 最小来产生容限 ξ，ξ 一旦产生就可以在容限 (X,ξ) 中通过求连通分支来找到某元素所属的类。有关学习问题后面进一步讨论。

二、诱导容限空间及其应用

定义 6 在容限空间 (X,ξ) 上，$A\subseteq X$，定义：

毛国君 硕士，讲师，主要研究兴趣为人工智能和神经网络

$\xi A=\{x \mid x\in X$ 且 $\exists\ a\in A,(a,x)\in\xi\}$.

定义 7 在容限空间 (X,ξ) 上，对 $X_1,X_2\subseteq X$，若存在正整数 n 使得 $\xi^n X_1\supseteq X_2$，则称 X_2 可由 X_1 生成，或者称 X_1 可生成 X_2。

显然如上定义的生成具有反身性和传递性，利用它可以定义下面的诱导空间。

定义 8 令 L_X 是集合 X 上所有子集的集合，$(L_X,\tilde{\xi})$ 称为 (X,ξ) 的诱导容限空间是指，对任意 X_1，$X_2\in L_X$，$(X_1,X_2)\in\tilde{\xi}$ 当且仅当 X_1 可由 X_2 生成，且 X_2 可由 X_1 生成。

我们再来看分类问题，通过输入 x，如果能得到它的连通分量 C(x)，而且正好找到某个类 C_i 使它们在诱导容限空间 $(L_X,\tilde{\xi})$ 中满足 $(C(x),C_i)\in\tilde{\xi}$，这样的分类系统就是成功的。即使系统不是这样精确，能保证存在某 C_i 使得 $C(x)\supseteq C_i$，也可以断定 $x\in C_i$，这是因为 $C_i(i=1,2,\cdots,M)$ 是一组互不相交的子集。然而这种处理方法在联想或整体识别问题上似乎遇到了麻烦，这是因为在事物的联想或整体识别中往往是给出部分信息而推知整个事物。实际上，给出若干部分信息 $x_1,x_2,\cdots,x_n$，都可以得到对应的连通分量 $C(x_1),C(x_2),\cdots,C(x_n)$，如果某一已知事物 $A\subseteq C(x_1)\cap C(x_2)\cap\cdots\cap C(x_n)$，那么 A 就可能是期望的结果。当然可能发生多个这样的已知事物 $A_1,A_2,\cdots,A_m$，使得 $A_i\subseteq C(x_1)\cap C(x_2)\cap\cdots\cap C(x_n)(i=1,2\cdots,m)$。这时可以认为部分信息量不足而宣告无法处理，也可以使用一定办法从这些事物中选一个最优的。

三、集合上的多容限及其学习算法

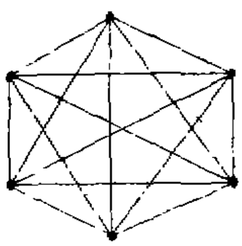

图 3

我们知道，一般的人工神经网络是交叉互联的拓扑结构。图 3 是一个全连型拓扑网络示意图，对于它的每个节点、若干的输入都必须产生对应输出，这一过程也可以看作是容限作用的结果。如果我们从整体上看，可以把所有神经元（节点）的功能汇集看作是整个对象上的容限，理论上说这个容限应该具有上面描述的性质。但是这种处理似乎也只限于理论描述，对实际应用指导意义不大。我们应该承认一个观察对象具有多容限的情况，并在此基础上探求其规律性。下面对这个问题做了初步尝试。

定义 9 设集合 X 具有多个容限 $\xi_1,\xi_2,\cdots,\xi_k$，$X_1,X_2,\cdots,X_{k+1}\subseteq X$，如果存在 k 个非负整数 $n_1,n_2,\cdots,n_k$，使得 $X_1\xi_1^{n_1}\supseteq X_2$，$X_2\xi_2^{n_2}\supseteq X_3,\cdots,X_k\xi_k^{n_k}\supseteq X_{k+1}$，则称 X_{k+1} 可由 X_1 经过 $\xi_1,\xi_2,\cdots,\xi_k$ 生成。

显然定义 9 是定义 7 的合理推广，即当 k=1 时就是定义 7。

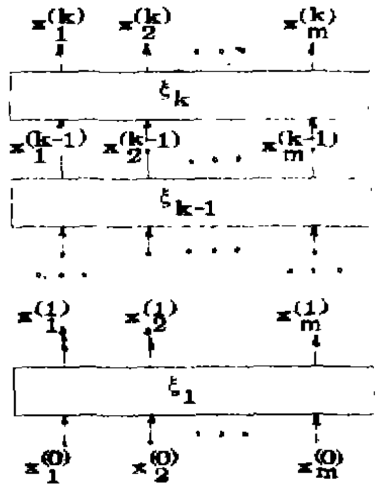

图 4

用定义 9 就可以更深入地讨论前向无反馈多层神经网络模型的学习问题了。图 4 给出了一个前向无反馈 k 层神经网络的黑箱式描述。输入是 $X_0=(x_1^{(0)},x_2^{(0)},\cdots,x_m^{(0)})$ 经过 $\xi_1,\xi_2,\cdots,\xi_k$ 产生中间输出 $X_1=(x_1^{(1)},x_2^{(1)},\cdots,x_m^{(1)})$，$X_2=(x_1^{(2)},x_2^{(2)},\cdots,x_m^{(2)}),\cdots$，最后输出 $X_k=(x_1^{(k)},x_2^{(k)},\cdots,x_m^{(k)})$。通过输入 X_0 经过 X_1，$X_2,\cdots,X_{k-1}$ 生成输出 X_k 的过程就是通过确定了的容限 $\xi_1,\xi_2,\cdots,\xi_k$ 作用的过程。而这些容限如何确定呢？这必须要靠人工神经网络的学习功能。这种模型可以采用递推学习算法。图 4 中的黑箱功能 $\xi_i(i=1,2,\cdots,k)$ 可用权重来描述。图 5 给出了 ξ_k 的内部结构示意图，其中 $w_{ij}^{(k)}(i=1,2,\cdots,m,j=1,2,\cdots,m)$ 表示第 i 个输入 $x_i^{(k-1)}$ 对第 j 个神经元 $u_j^{(k)}$ 的作用权重。首先给定预期输出向量 $(d_1,d_2,\cdots,d_m)$，通过它自上而下递推调整权重，直到满意为止。

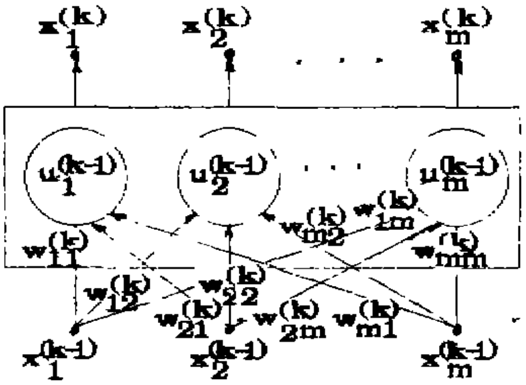

图 5

我们先看最上一层．为了推导方便，把功能函数简化为：

$$x_i^{(k)}=\sum_{j=1}^{m}w_{ij}^{(k)}\cdot x_j^{(k-1)}.$$

于是，输出误差的均方和 $E^{(k)}=\sum_{i=1}^{m}(d_i-x_i^{(k)})^2$。

令 $W^{(k)}=(w_{11}^{(k)},w_{12}^{(k)},\cdots,w_{1m}^{(k)},\cdots,w_{mm}^{(k)})$，则 $E^{(k)}$ 相对于 $W^{(k)}$ 的梯度为：

$$\nabla E^{(k)}=(\frac{\partial E^{(k)}}{\partial W_{11}^{(k)}},\frac{\partial E^{(k)}}{\partial W_{12}^{(k)}},\cdots,\frac{\partial E^{(k)}}{\partial W_{1m}^{(k)}},\cdots,\frac{\partial E^{(k)}}{\partial W_{mm}^{(k)}}),$$

若令 $\triangle W^{(k)}=(\triangle w_{11}^{(k)},\triangle w_{12}^{(k)},\cdots,\triangle w_{1m}^{(k)},\cdots,\triangle w_{mm}^{(k)})$，则

$$\triangle E^{(k)}=\nabla E^{(k)}\cdot\triangle W^{(k)}=\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\partial E^{(k)}}{\partial w_{ij}^{(k)}}\cdot\triangle w_{ij}^{(k)}.$$

如果能使 $E^{(k)}$ 不断减少且收敛，那么就可以通过重复学习训练找到合适的 $W^{(k)}$，把 ε_k 定下来。显然，欲使 $\triangle E^{(k)}\leqslant 0$，可取

$$\triangle w_{ij}^{(k)}=-\alpha\frac{\partial E^{(k)}}{\partial w_{ij}^{(k)}}(\alpha>0)。$$

由 $E^{(k)}=\sum_{i=1}^{m}(d_i-x_i^{(k)})^2=\sum_{i=1}^{m}(d_i-\sum_{j=1}^{m}w_{ij}^{(k)}\cdot x_j^{(k-1)})^2$

容易得到

$$\frac{\partial E^{(k)}}{\partial w_{ij}^{(k)}}=-2(d_i-x_i^{(k)})\cdot x_j^{(k-1)},$$

故

$$\triangle w_{ij}^{(k)}=2\alpha(d_i-x_i^{(k)})\cdot x_j^{(k-1)}=\alpha_k(d_i-x_i^{(k)})\cdot x_j^{(k-1)}$$

其中 $\alpha_k=2\alpha$。

有了 $\triangle w_{ij}^{(k)}$ 就可以产生新的 $w_{ij}^{(k)*}$，即 $w_{ij}^{(k)*}=w_{ij}^{(k)}+\triangle w_{ij}^{(k)}$。理论上讲，由于新 $w_{ij}^{(k)*}$ 的产生使得我们可以通过方程组 $d_i=\sum_{j=1}^{m}w_{ij}^{(k)*}\cdot d_j^{(k-1)}(i=1,2,\cdots,m)$ 产生第 $k-1$ 层的预期输出 $(d_1^{(k-1)},d_2^{(k-1)},\cdots,d_m^{(k-1)})$，再通过上述方法又得到第 $k-1$ 层新的权重 $W^{(k-1)*}$，如此下去便可完成一趟学习训练。但是，注意到 $d_i=\sum_{j=1}^{m}w_{ij}^{(k)*}\cdot d_j^{(K-1)}$ 的解法的复杂性，可以利用上一层的部分结果来产生下一层的权重以调整量 $\triangle w_{ij}$，不断尝试直到成功为止．下面我们给出一个前向无反馈多层神经网络的学习算法．

前向无反馈多层模型的递推算法描述为：

```
begin
对所有 i,j=1,2,…,m,置 w_ij 一个初始值;
repeat
  输入 x_1^(0),x_2^(0),…,x_m^(0);
  对所有 i=1,2,…,m do x_i^(1) = Σ_{j=1}^{m} w_ij^(1) · x_j^(0);
  对所有 i=1,2,…,m do x_i^(2) = Σ_{j=1}^{m} w_ij^(2) · x_j^(1)
  ……
  对所有 i=1,2,…,m do x_i^(k) = Σ_{j=1}^{m} w_ij^(k) · x_j^(k-1);
  输入一个预期输出 d_1,d_2,…,d_m;
  if E = Σ_{i=1}^{m} (d_i - x_i^(k))^2 < 足够小的 ε then 结束;
  输入 α_k;
  对所有 i=1,2,…,m do
  begin
    置 δ_i^(k) = α_k(d_i - x_i^(k));
    对所有 j=1,2,…,m do
    begin
      置△w_ij^(k) = δ_i^(k) · x_j^(k-1); 置 w_ij^(k) = w_ij^(k) + △w_ij^(k)
    end
  end;
  输入 α_{k-1};
  对所有 i=1,2,…,m do
  begin
    置 δ_i^(k-1) = α_{k-1}(Σ_{p=1}^{m} δ_p^(k) · w_pi^(k));
    对所有 j=1,2,…,m do
    begin
      置△w_ij^(k-1) = δ_i^(k-1) · x_j^(k-2);
      置 w_ij^(k-1) = w_ij^(k-1) + △w_ij^(k-1)
    end
  end;
  ……
  输入 α_1;
  对所有 i=1,2,…,m do
  begin
    置 δ_i^(1) = α_1(Σ_{p=1}^{m} δ_p^(2) · w_pi^(2));
    对所有 j=1,2,…,m do
    begin
      置△w_ij^(1) = δ_i^(1) · x_j^(0);
      对所有 j=1,2,…,m do
      begin
        置△w_ij^(1) = δ_i^(1) · x_j^(0); 置 w_ij^(1) = w_ij^(1) + △w_ij^(1)
      end
    end
  end
end;{repeat 结束}
```

上面我们对容限空间理论进行了扩展，并通过它重新认识了人工神经网络的有关问题。从而可以看出容限空间理论的完善对于神经网络计算机和人工智能以及模式识别等问题的研究有着重要的指导意义。

神经网络 Hopfield网络 学习问题

计算机科学 1994Vol. 21№. 5

30-34

非对称型 Hopfield 神经网络的学习问题*)

吴福朝 张 铃
（安徽大学计算机系 合肥 230039）

TP18

摘 要 In this paper, the learning problem of non-symetrical Hopfield neural networks is turned into solving unequalites and planning problem. Then famous ellipsoid method and planning method in mathematics have been applied to satisfactorily solve it.

关键词 Neural network、Ellipsoid method, Karmarkar method, Learing algorithm.

1. 引 言

联想记忆是人工神经元网络的重要功能之一，Hopfield 网络是一种重要的应用于联想记忆型网络。为了实现记忆功能，我们总希望通过训练使样本成为网络的稳定状态。然而，Hopfield 网络利用 Hebb 规则训练，求得的连接权矩阵 W：

$$w_{ij}=\begin{cases}\sum_{\alpha=1}^{M} x_i^\alpha x_j^\alpha, & i\neq j\\ 0, & i=j\end{cases}$$

进行联想记忆时，稳定状态出现的规律并非如此简单，主要表现在[1]：

(a)给定一组样本，经过学习（训练），网络将有一组稳定状态，但是给定的样本可能不是稳定状态。即使给定的样本均成为稳定状态、网络有可能出现很多其他多余的稳定状态（称之为伪稳定状态），这些伪稳定状态会极大地损害联想记忆功能。

通过其他方式进行学习的任何对称型（即权矩阵为对称阵）Hopfield 网络，也会出现上述问题。因此有必要引进非对称型 Hopfield 网络[2]。文[2]的研究表明，非对称型 Hopfield 网络，较之对称型具有良好的记忆联想功能，但是它的结构却要复杂得多，这主要表现在：

(b)非对称型网络的稳定状态（吸引子）、有可能不是网络能量函数的极小点；而对称型网络在 $w_{ii}=0, \forall i$ 的情况下，所有稳定状态均为能量函数的极小点。

本文主要针对问题(a)和(b)，来讨论非对称网络的学习问题。即给定样本组：

$$\mathscr{X}=\{x^1,x^2,\cdots,x^M\}\subset\{-1,1\}^N$$

其中 N 为自然数，是训练样本 x^α 的维数，$1\leqslant\alpha\leqslant M$，求连接权矩阵 W^* 使之满足条件：

(Ⅰ) $\mathscr{X}$ 为相应网络的一组稳定状态；

(Ⅱ) $\mathscr{X}$ 中每一样本是相应网络能量函数的极小点；

(Ⅲ)相应网络的伪稳定状态要尽可能的少；

为了使相应网络有良好的联想功能，还要求满足条件：

(Ⅳ) $\mathscr{X}$ 中每一点的吸引半径>1。

无论是对称型还是非对称型，网络状态演化运行规律通常有同步运行和异步运行两种方式。

同步运行时，在每一单位时间中，网络的所有神

*)国家基础研究项目攀登计划基金资助。吴福朝 副教授。张铃 教授。

参考文献

[1]Zmoeeman, E. C., The Topology of the Brain and Visual Percaption. In Topology of 3—manifolds and Related Topics, Prentice-Hall, INC., Englewood Cliffs N. J., 1962

[2]陈霖，格式塔和容限空间，心理学报 1984.3

[3]杨行峻等，人工神经网络，高等教育出版社，1992，9

[4]王汝笠等，第六代计算机——人工神经网络计算机，科学技术文献出版社，1992，12

[5] Mao Guojun., Lattice-valued Logic, Intelligent Mathematics and Their Applications. Proc. IEEE ISMVL—89, 1989

[6]夏道行等，线性拓扑空间引论，上海科学技术出版社，1986，11