计算机科学1993Vol.20№.5

预测结果的一种计算方法*)

张师超 （国家智能计算机研究开发中心 北京100080）
（广西师范大学数学系 桂林541004）

摘 要

This paper presents a computing method of forecasted result in [1] by cited relative number formula in statistics. And this method is applied to knowledge automated acquisition.

1. 引言

预测是人类专家经验知识在生产实践中的综合运用，是决策的依据。决策的好坏可决定一个厂矿企业或者一个国家的前途和命运，因此，预测有着重要的实际意义。

目前，多数预测的建模是基于概率学的，硬性地将预测问题归为简明严谨的数学模型，尽管在某些特定条件下可行，但其实效性和可行性也难以保证。它有如下弱点：

·预测时使用方程的有效性依赖于对这些方程的扩展假设是否正确。例如，正态分布随机变量，由这些假设所产生的偏差效果往往不大清楚。

·方程自身有疑问。这也许是因为不能从理论上提供一类特殊函数，因而往往假定函数是线性的。

·缺乏做详细预测的数据，或数据本身有错误。

可见，预测的研究仍处于理论探索和实践中。我们的观点是，有机结合计算机科学和概率统计学中的成熟结果，可以构造更为有效的预测模型。

本文在文[1，3，4]的基础上，运用概率学中的线性相关公式，建立预测结果的一种估计模型，并将这一方法运用到知识的自动获取中，得到了一个知识自动整理的模型。

*) 国家863计划资助课题（863-306-601-132）。

为了简化描述，本文采用文[1，3，4]中的一些记号和规定。

2. 预测结果的估计方法

文[1]讨论历史数据的前兆依赖性，给出如下预测模型。

定义2.1 设$U=A_1A_2\cdots A_k$是一个属性集合，$T\subseteq U$是时序集合，则U上的数据库H称为历史数据库。设查询历史$t_1t_2\cdots t_{n+1}$得到$\bar{u}=t_1t_2\cdots t_n$与目前发生的事实$\bar{v}=t_1't_2'\cdots t_n'$十分接近，即对某个充分小的$\varepsilon>0$有$\|t_i-t_i'\|<\varepsilon$，$i=1,2,\cdots,n$。那么由领域内专家的经验知识可知，对即将发生的事件t'_{n+1}，存在一个充分小的$\delta>0$，使得$\|t_{n+1}-t'_{n+1}\|<\delta$，即，可由$t_{n+1}$预报$t'_{n+1}$。这种数据的依赖称为前兆依赖。

在实践运用中，与目前事实$\bar{v}$十分相似的历史$\bar{u}$往往是不唯一的。因此，我们必须要对这些历史作适当的处理才能做出预报的结果。下面，我们采用线性相关公式来确定预报的结果。

设目前发生的事实为$\bar{v}=t_1't_2'\cdots t_n'$，而即将发生的事件为$t'_{n+1}$，与$\bar{v}$十分相似的历史区间为m个：$\bar{u}_i=t_{i1}t_{i2}\cdots t_{in}$，及$t_{i(n+1)}$，$i=1,2,\cdots,m$。并且，我们假设$t_{1(n+1)}$，$t_{2(n+1)}$，$\cdots$，$t_{m(n+1)}$十分相似（若有x个例外事件，则若$(m-n)/m\geq\lambda$，$\lambda$为专家给定

的阈值，亦可用这种方法处理，但可信度要做适当的考虑）。

将 $\bar{v}$ 和 $\bar{u}_i$ 中的数据的模糊词数值化后得：

$\bar{u}_i=(c_{i1},b_{i1}),(c_{i2},b_{i2}),\cdots,(c_{i(n+1)},b_{i(n+1)})$，$i=0,1,2,\cdots,m$，$\bar{u}_0=\bar{v}$。

根据假设 $\bar{u}_i$ 与 v 十分相似，即 $\|t_{ij}-t_j'\|=\|c_{ij}-c_{0j}\|+\|b_{ij}-b_{0j}\|<\varepsilon$，$\varepsilon>0$ 为某个充分小的数。于是有：

$\|c_{ij}-c_{0j}\|=0$，$\|b_{ij}-b_{0j}\|<\varepsilon$

因此，$c_{ij}=c_{0j}$，用 c_j 表示 c_{ij}，则上述表示可化成：

$\bar{u}_i=(c_1,b_{i1}),(c_2,b_{i2}),\cdots,(c_{n+1},b_{i(n+1)})$。

根据前兆依赖性可知，(c_1,b_{i1})，(c_2,b_{i2})，…，(c_n,b_{in}) 与 $(c_{n+1},b_{i(n+1)})$ 间存在一些联系。由于 $c_1,c_2,\cdots,c_{n+1}$ 均不变，所以，只有 $b_{i1},b_{i2},\cdots,b_{i(n+1)}$ 在变化，并通常假设这种联系是线性的。我们分别求 b_{ij} 与 $b_{i(n+1)}$ 间的相关系数，$j=1,2,\cdots,m$。

令 $b_m^{(j)}{}_n=\mathrm{Min}\{b_{1j},b_{2j},\cdots,b_{mj}\}$，

$b_m^{(j)}{}_x=\mathrm{Max}\{b_{1j},b_{2j},\cdots,b_{mj}\}$，

$j=1,2,\cdots,m$。

取 $b_j=(b_m^{(j)}{}_n+b_m^{(j)}{}_x)/2$，

$\theta_j=(b_m^{(j)}{}_x-b_m^{(j)}{}_n)/2$。

现在先定义两组数的运算，设数据为：

x	x_1	x_2	…	x_m
y	y_1	y_2	…	y_m

记 $\sum x_iy_i=x_1y_1+x_2y_2+\cdots+x_my_m$，

$\sum x_i=x_1+x_2+\cdots+x_m$，

$\sum y_i=y_1+y_2+\cdots+y_m$，

$\sum x_i^2=x_1^2+x_2^2+\cdots+x_m^2$。

根据概率学中相关系数公式我们可求出 (c_j,b_{ij}) 与 $(c_{n+1},b_{i(n+1)})$ 间的相关系数 r_j：

$$r_j=\frac{\left[m\sum_{i=1}^{m}(b_{i(n+1)}\cdot(b_{ij}-(b_j-b_{n+1})))-(\sum_{i=1}^{m}(b_{ij}-(b_j-b_{n+1})))\sum_{i=1}^{m}b_{i(n+1)}\right]}{\sqrt{\left[m\sum_{i=1}^{m}(b_{ij}-(b_j-b_{n+1}))^2-(\sum_{i=1}^{m}(b_{ij}-(b_j-b_{n+1})))^2\right]\left[m\sum_{i=1}^{m}b^2{}_{i(n+1)}-(\sum_{i=1}^{m}b_{i(n+1)})^2\right]}} \tag{1}$$

一般，r_j 越大，则 (c_j,b_{ij}) 与 $(c_{n+1},b_{i(n+1)})$ 就越相关。对 r_j 值的要求是随 m 的大小来决定的。对于给定的 m，设临界值为 $r^{(m)}$，则若 $r_j<r^{(m)}$，我们称 (c_j,b_{ij}) 与 $(c_{n+1},b_{i(n+1)})$ 无关。

有了 r_j 后，我们可以计算出每个前兆的权重值 w_j：

$$w_j=r_j/(r_1+r_2+\cdots+r_n) \tag{2}$$

$j=1,2,\cdots,n$。

现在，我们来给出计算预测结果的公式。前面已统计出每个前兆的中点值 $b_j(j=1,2,\cdots,n)$，我们可以得到现有迹象的 t' 与 (c_j,b_j) 的偏差的加权和为：

$$b=\sum_{j=1}^{n}w_j(b_{0j}-b_j)。 \tag{3}$$

为了计算预测结果，我们需要计算出前兆的最大偏差量 b_{max} 和最小偏差量 b_{min}：

$$b_{max}=\sum_{j=1}^{n}w_j\cdot\theta_j \tag{4}$$

$$b_{min}=\sum_{j=1}^{n}w_j\cdot(-\theta_j)=-\sum_{j=1}^{n}w_j\cdot\theta_j \tag{5}$$

由相关系数和权重的确定方式，我们近似假定预测结果的非确定值与前兆的加权偏差 b 成正比。于是，设 x 表示前兆加权改变量，y 表示预测结果的确定值改变量，则 x 和 y 满足：

$$y=\frac{2\theta_{n+1}}{b_{max}-b_{min}}\cdot x-\frac{\theta_{n+1}(b_{max}+b_{min})}{b_{max}-b_{min}} \tag{6}$$

于是，预测结果的确定值为：

$$b_{0(n+1)}=b_{n+1}+\frac{2\theta_{n+1}}{b_{max}-b_{min}}\cdot b$$

$$-\frac{\theta_{n+1}(b_{max}+b_{min})}{b_{max}-b_{min}}$$

$$=b_{n+1}+\frac{2\theta_{n+1}\cdot b-(b_{max}+b_{min})}{b_{max}-b_{min}} \quad (7)$$

下面，我们举例来说明上述公式的使用。

设历史数据为：

$(c_1, 0.7)$，$(c_2, 0.12)$，$(c_3, 0.5)$；

$(c_1, 0.8)$，$(c_2, 0.2)$，$(c_3, 0.6)$；

$(c_1, 0.9)$，$(c_2, 0.28)$，$(c_3, 0.7)$；

$(c_1, 0.74)$，$(c_2, 0.16)$，$(c_3, 0.55)$；

$(c_1, 0.86)$，$(c_2, 0.24)$，$(c_3, 0.65)$。

现在发生的迹象为：

$(c_1, 0.85)$，$(c_2, 0.21)$。

根据前面的公式我们可得：

$b_{min}^{(1)}=0.7$，$b_{min}^{(2)}=0.12$，$b_{min}^{(3)}=0.5$，

$b_{max}^{(1)}=0.9$，$b_{max}^{(2)}=0.28$，$b_{max}^{(3)}=0.7$，

$b_1=0.8$，$b_2=0.2$，$b_3=0.6$，

$\theta_1=0.1$，$\theta_2=0.08$，$\theta_3=0.1$，

由（1）式可求得：

$r_1=0.9970548$，$r_2=1$

由（2）可得：

$w_1=r_1/(r_1+r_2)=0.4992626$，

$w_2=0.5007374$

由（3）式可得：

$$b=0.4992626\times(0.85-0.8)+0.5007374\times(0.21-0.2)$$

$$=0.0299704$$

由（4）和（5）得

$$b_{max}=0.4992626\times0.1+0.5007374\times0.08$$

$$=0.0899851$$

$$b_{min}=-0.0899851$$

由（7）式可得

$$b_{03}=b_3+\frac{2\theta_3\cdot b-(b_{max}+b_{min})}{b_{max}-b_{min}}$$

$$=0.6+\frac{2\times0.1\times0.0299704}{0.0899851+0.0899851}$$

$$=0.6333055$$

即预测结果应为（c_3，0.6333055）。

3．相关系数在知识自动获取中的应用

知识获取是人工智能中的瓶颈问题，这是由多方面的因素引起的。这里，我们用相关系数建立规则前件的权重求解公式，以及用相关系数消除多余的前件的方法。它为知识的自动获取提供了一种有效的自动检测手段。

因为专家的知识是从实践中获取的，即从大量的实际数据中提炼出来，并经过多次的实际运用和检验而形成的。我们把这些实践中的数据称为该知识的数据源。

人类专家从数据源中获取知识这一过程可以由计算机来模拟。用计算机获取规则形式表示的知识完全可以根据上一节提供的方法来确定规则的每个前提的权重，并根据相关系数值来确定删除规则的无用前提。

计算规则的每个前提的权重方法是，将第二节的m个历史数据看成是数据源，或者将数据源中的数据按上一节处理历史数据的方式进行处理，求出每个前提与结果的相关系数，按（2）式求每个前提的权重，

由于数据源中的数据是粗糙的，数据的收集也不是为获取某条知识而采集的，因此，从数据源中获取知识时，必须能将无用的或者与知识无关的部分去掉，以便保证知识的实质性质。例如，设$A\wedge B\wedge C\rightarrow D$是获取的知识，而C与D是无关的。在推理过程中若已知A和B为真，要推出D为真还需等待C为真。象C这种前提条件应从该规则中去掉。确定规则的前提条件与结果的关系同样可使用确定规则的每个前提条件的权重值的方式求出每个前提条件与结果的相关系数r，若$r\geqslant r_0$，则该前提条件与结果相关，应被保留在规则中；若$r<r_0$，则该前提条件与结果无关，将它从该规则中删掉，r_0是专家给定的一个临界值。r_0的大小通常与数据源中数据的个数有关，它的取法在概率方面的书中有所讨论。

模糊用系统 神经网络 (10)

计算机科学1993Vol.20№.5

51~54

模糊系统与神经网络的融合

TP18

周长久 （大连铁道学院 大连116022）

摘 要

In this paper, based on the analysis of contact point of Fuzzy System(FS) and Neural Neteworks(NN), several fusion methods of FS and NN are proposed. Problems of the fusion technology are studied. Applications of the fusion technology are showed to emphasize its advantages. Finally, some problems to be solved are proposed, and the future prospects are described.

1. 引言

近年来，模糊系统FS(Fuzzy System)与神经网络NN (Neural Network) 的研究取得了很大进展。FS的显著特点是它能直接地表示逻辑，适于直接的或高一级的知识表达；而NN则是通过学习功能获得用数据表达的知识，这种知识在NN中是隐含表达的，难以对其进行分析。可以这样说，FS比较适于自顶向下过程，而NN则更适于自低向上过程。随着研究范围的扩大，逐渐改变了上述两个领域之间相互独立的关系，即一个领域的固有缺乏可以通过另一个领域的优点来补偿，这就导致了一个新的领域的产生，即FS与NN的融合 (fusion) 技术。

最早将模糊思想引入到神经网络的论文是由S.C.Lee[1]发表的。但直到近几年FS与NN两个领域飞速发展，人们才开始特别关注FS与NN的关系。例如，Kosko[2]建议在专家系统推理中把NN与FS结合起来，Shiue[3]和Keller[4]利用模糊集理论来设计NN的学习算法，Keller[5]还提出了一种能完成模糊推理的NN结构，C.T.Lin[6]研究了基于NN的模糊控制器。其中，大多数研究是将NN的学习功能融合到FS中，而将FS渗透到NN中的研究却很少[7]。

本文在分析FS与NN关系的基础上，回顾了融合技术的发展，并重点研究了FS与NN融合的主要途径，最后，展望了一些可能的应用并提出一些尚需进一步探讨的问题。

2. FS与NN的联系

设神经元有n个输入x_1，x_2，…，x_n，则其输出为：

$$y=f(\sum w_i x_i-\theta) \qquad (1)$$

其中f为阈值函数。

对于一个有n个输入的模糊逻辑系统，其l条规则为：（$i=1,2,\cdots,L$）

IF x_1 is A_{1i} AND…x_n is A_{ni},

THEN y is B_i

周长久 硕士，讲师，副主任，主要研究方向为模糊系统及控制、神经网络和专家系统等。

参 考 文 献

〔1〕唐常杰等，历史数据的前兆依赖及其数学性质，中国科学（A辑），2（1989）

〔2〕史忠植，智能决策，国家智能计算机研究开发中心

〔3〕张师超，一个预测模型及其性质，计算机学报，4（1992）

〔4〕张师超，前兆依赖的取值及数据划分，计算机学报，3（1993）