基于构造性覆盖算法的离群数据挖掘研究*)

张 旻[1,2] 张 铃[2]

(解放军电子工程学院204研究室)[1] (国家教育部安徽大学人工智能与信号处理重点实验室 合肥230039)[2]

摘 要 本文提出一种通过构造覆盖领域进行离群点(outlier)挖掘的新方法。由于覆盖领域构造的特殊性,使得覆盖算法非常适合离群点的挖掘。在分析覆盖模型的基础上,给出了覆盖模型的离群点的定义和算法步骤。这样将复杂的离群点挖掘问题变成十分简单的覆盖领域样本分析问题,而且算法十分直观,并能很好地解释离群点的含义,同时适合对高维及海量数据的处理。本文给出实验例子,结果表明该方法是有效可行的。

关键词 数据挖掘,离群点,覆盖算法,覆盖样本

The Research of Outlier Mining Based on Constructed Covering Algorithm

ZHANG Min[1,2] ZHANG Lin[2]

(204 Research Room of Electronic Engineering Institute of PLA)[1]

(Key Lab of IC&SP at Anhui University, Ministy of Education, Hefei, 230039)[2]

Abstract A new algorithm of outlier mining based on the covering algorithm is proposed in this paper. Since the particularity of constructing the covering domain, the covering is very suitable for the outlier mining. After analyzing the covering model, the definition of the outlier and procedures of algorithm according the covering model is given. Through this way, the complicate problem of outlier mining is changed into very simple analyzing covering samples problem, which is intuition and give good explanation of the outlier. And the algorithm is suitable for the high dimension and huge mount of data processing. Some experiments are given, and the results show the effective of the algorithm.

Keywords Data mining, Outlier, Covering algorithm, Covering samples

1 引言

KDD(knowledge discovery in databases)是从大量数据中发现正确的新颖的潜在有用并能够被理解的知识的过程[1]。现有的KDD研究大多集中在发现适用于大部分数据的常规模式。但在一些应用中,如电子商务和金融服务领域中的欺诈等犯罪行为监测,有关例外情况的信息比常规模式更有价值。目前,这样的研究正得到越来越多的重视。离群数据已在统计学领域得到广泛研究[2],但基于统计的方法需要用户建立数据点的概率分布模型,应用时需事先知道数据集的分布和分布参数等信息。Knorr 和 Ng[3]提出基于距离的离群数据挖掘方法,但这种方法中的距离难以确定,而且没有离群数据的离群衡量测度。Arning 等[4]提出一种基于偏离的离群数据发现方法,需要确定相异函数进行离群数据挖掘,但若其相异函数的选取不合适,则得不到满意的结果。高维数据的特性完全不同于低维数据,因此高维空间中的离群点发现方法势必不同于传统的离群点发现方法。而目前还没有一个十分有效的高维空间离群点的挖掘方法。

文[5,6]提出一种 M-P 模型的几何表示,在此基础上形成了神经网络覆盖算法。算法首先将 n 维样本经过归一化处理投影变换到 $n+1$ 维超半球上,因此在超球上位置点蕴涵着样本内在信息。在构造覆盖网络时充分利用了样本投影点在超球面上的位置信息,而位置信息反映了样本自身所有属性的特性,因而覆盖领域内的样本就是一种值得挖掘的模式。本文首次提出一种新颖的基于覆盖算法的离群数据挖掘算法,该算法根据覆盖领域的样本数来挖掘离散样本点,非常直观可信。算法在构造覆盖网络时是由数据自身特性构造而成,且适合海量和高维数据的运算[8~10],因此基于覆盖模型而进行的离群点挖掘同样适合海量和高维数据的处理。

为了保持一定的完整性,本文首先简单介绍文[5]中给出的覆盖算法几何表示,然后分析利用覆盖模型进行离群点挖掘的合理有效性,给出覆盖模型的离群点定义,挖掘离群点的具体步骤,最后给出实验和结论。

2 构造性学习方法——覆盖算法简介

1943年 McCulloch 和 Pitts[11]根据神经元传递中的"0,1率"和神经传递中信号不但有不同的强度,而且有兴奋和抑制两种情况,第一次提出来神经元的数学模型(M-P 模型),此模型一直沿用至今。

将神经元看出是一个有 n 个输入和一个输出的元件,此元件的激励函数可表示为 $y=\sigma(W^* x-\theta)$,其中 σ 是符号函数,即:

$$\sigma(x)=\begin{cases}1, & x>0\\ -1, & x\leqslant 0\end{cases}$$

现在分析一下这个模型的几何意义。

2.1 超平面表示

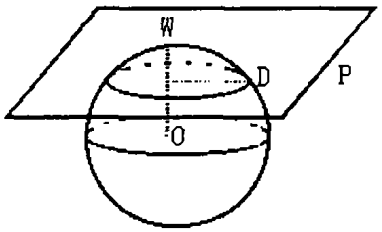
由神经元激励函数的定义可知,它由两个函数复合而成的,其中第1个函数为 $(W*x-\theta)$,若令其等于零,得 $W*x-\theta=0$。这个方程在 n 维空间中表示一个超平面 P,当 $(W*x-\theta)>0$ 时,表示点落在超平面的正半空间内,此时,$\sigma(W*x-\theta)=1$,当 $(W*x-\theta)<0$ 时,表示点落在超平面的负半空间内,此时 $\sigma(W*x-\theta)=-1$。于是,一个 M-P 神经元的功能可看成是一个由超平面划分的空间位置的识别器。这给神经元一个相当直观的解释。人们也曾利用这个直观的理解研究神经网络的学习问题。当空间存在两三个平面时,尚可直观理解,而当 n,m 较大时,即 n 维空间中 m 个超平面的相交情况时,就非常复杂、很不直观。总之,很难用超平面的模型帮助我

*)国家自然科学基金科研基金(60175018)。张 旻 副教授,博士生,研究方向:信号处理、机器学习、人工智能等。张 铃 教授,博士生导师,研究方向:智能计算,人工智能等。

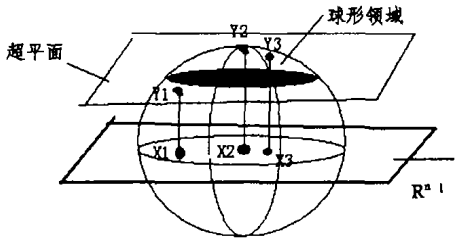
们直观地理解神经网络的内在性质。故此后很少有人再用神经元的超平面表示的几何意义来帮助研究神经网络的学习问题了。

2.2 超球面上的"领域"表示

几何直观往往是研究的很好的向导，超平面的方法遇到困难，但是张铃教授对此进行了改变[5]，将会给我们提供新的直观帮助。



(a) 超平面P与超球面相交，形成"球形领域"的示意图



(b) 从D→S^n变换图

图1

如图1(a)，若我们限定输入向量的长度相等，即输入向量限定在n维空间的某个球面上，那么这时$(W*x-\theta)>0$就表示球面上落在p正半空间的部分，这个部分恰好是球面上的某个"球形领域"，若取W与x等长，则这个"球形领域"的中心恰好是W，其半径为$r(\theta)$，它是θ的单调下降的函数。若我们取$\sigma'(x)=\begin{cases}1, x>0\\0, x\leqslant 0\end{cases}$，且取神经元的激励函数为$\sigma'(W*x-\theta)$，则一个神经元的激励函数正好是它所代表的球面上"球形领域"的特征函数。这样，我们就能非常直观地进行神经网络的各项研究。

通过这种变换，我们就将神经网络的最优设计问题转化成某种最优覆盖问题。当给定的输入向量的长度不相等时，可用下面给出的方法将它变换成长度相等的情况。设输入的定义域为n维空间中的有界集合D，令S^n是$n+1$维空间中的n维的超球面，作变换

$$T: D\to S^n, T(x)=(x, \sqrt{r^2-\|x\|^2})$$

将样本点映射到球面S^n上，其中$r\geqslant\max\{|x_i|\}$。

这个变换几何直观上可以理解为，将D看成是位于$n+1$维空间中过原点的一个n维超平面上，而且D位于要S^n的内部。则变换T就是将D上的点垂直投射到S^n的上半球面上(图1(b))。这种变换显然是一一对应的。如上面所述，这时每一个神经元(W,θ)，就是在超球面S^n上以W为中心，以$r(\theta)$为半径的一个"球形领域"的特征函数(其中$r(\theta)=r(\cos(\theta|R))$)。

2.3 覆盖领域的构造

设样本集S为$S=\{(x^t=(x^t,y^t),t=0,1,\cdots,p-1\}$，集合中不同的$y^t$(类标记)只有$k$个，不妨设$y^t$的头$k$个的值均不相同。令样本的输出为$y^t$的样本标记的集合为$I(t)$(即$I(t)=\{i|y^i=y^t\}$)，其对应的输入集合记为$P(t)$，$t=0,1,2,\cdots,k-1$。并将输入样本的上标按$I(0),I(1),\cdots,I(k-1)$的顺序排列，于是，我们能够取一批"球形领域"$\{\{C_j^t, t=0,1,\cdots,k-1; j=1,2,\cdots,k_t\}\}$。令$C=\bigcup C_j^t$，使得$C^t$只覆盖$j$属于$I(t)$的$x^t$，而不覆盖$j$不属于$I(t)$的$x^t$，且$C^t$互不相交。这样球形领域就能达到学习分类的目的。固定t，令$P(t)=\{x^i|i\in I(t)\}$，对样本输入$x^t\in P(t)$，覆盖领域按式(1)构造。

$$\begin{aligned}d^1(i)&=\max_{j\in I(t)}\{\langle x^i,x^j\rangle\}\\d^2(i)&=\min_{j\in I(t)}\left\{\begin{matrix}\langle x^i,x^j\rangle|\langle x^i,x^j\rangle\\>d^1(i)\end{matrix}\right\}\\d(i)&=\frac{1}{2}(d^1(i)+d^2(i))\end{aligned}\qquad(1)$$

其中$\langle x,y\rangle$表示内积。

计算$d(i)$，作以x^i为中心、阈值$\theta=d(i)$的覆盖C_j^i，并通过求球形领域的重心和平移领域中心位置，使之可以覆盖更多的样本点，并按此方法求出样本的全部覆盖领域。

从式(1)中可以看出在构造覆盖领域时充分利用了样本的类标记和样本的相似性这两个重要特性，在覆盖样本时尽可能多的将同类样本覆盖进来。对于覆盖不住的样本，再产生新的覆盖领域继续覆盖，如此反复，直至全部覆盖。显然同一覆盖领域的样本具有很好的相似性，因为它们都投影在超球面上相近的位置，容易聚集在同一领域。对于远离同类点的离群样本，由于投影点离同类的投影点较远，因此形成了孤点，很难进入其它的样本覆盖领域，只能由新的覆盖领域进行覆盖，这就为离群点的分析创造了条件。

3 覆盖算法的离群点挖掘

3.1 覆盖领域的分析

为了形象说明覆盖算法问题，图2从二维覆盖图进行说明。

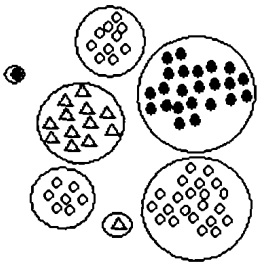


图2 原样本覆盖图

样本共分为黑实心点、空心圆点、空心三角三类，分别有2、3、2个覆盖领域，每个领域的覆盖样本数不同。覆盖领域内的样本有以下特点：一是同一领域内的样本具有相同的类标记。二是同一领域内样本具有很强的相似性。三是同一类标记的样本，如果差异较大，不会在同一覆盖领域内，会形成多个领域。四是相似的样本，如果类别不同，也不可能聚集在同一覆盖领域。显然覆盖网络中的每个覆盖领域就是一个很好的模式，值得挖掘。

3.2 覆盖算法的离群点挖掘

到目前为止，离群点还没有一个正式的为人们普遍接受的定义。Hawkin的定义[7]揭示了离群点的本质："离群点的表现与其它点如此不同，不禁让怀疑它们是由另外一种完全不同的机制产生的。"

由于离群点的特性，使得离群点与正常的点在超球面的投影的位置会有很大的区别。因此，离群点就很难和正常点构造在同一覆盖领域，往往是独自形成覆盖领域，这样就使得通过覆盖学习后离散数据的挖掘变得十分简单，因为根据式(1)而构造的覆盖领域，如果领域中的覆盖样本是很少量的，一定

是样本的特性与常规的样本投影点的距离相差很远或者样本远离本类样本群而落入了其它类别的样本群中，形成了离群点。而如果一个覆盖领域的样本数量很多的点说明它们的原样本有很大的相似性，因此是正常样本点。这样就将一个十分复杂的孤点挖掘问题转变成十分简单的覆盖样本分析问题。因此采用对学习后形成覆盖领域的样本点分析来挖掘离群点，即十分直观，又十分合理，而且构造覆盖领域是根据样本自身特性，按式(1)构造而成，简单易行，且适合高维和海量数据的构造分析，克服了其它常规算法所需的各种判别条件等限制。下面我们给出覆盖算法的离群点定义。

定义1 C_j^i 为第 i 类样本的第 j 个覆盖领域，$card(C_j^i)$ 为该覆盖领域的样本数量，$card(C^i)$ 为该样本类别为 i 的样本数量，该覆盖领域的支持度 s_j^i 表示：

$$s_j^i = card(C_j^i)/card(C^i)$$

定义2 给定阈值 ε，如果 $card(C_j^i) \leqslant \varepsilon$，则 C_j^i 覆盖领域内的样本为离群点。

定义3 给定阈值 ε_1，如果支持度 s_j^i 满足条件 $s_j^i \leqslant \varepsilon_1$，则 C_j^i 覆盖领域内的样本为离群点。

实际使用时，当样本数量较少时，就可以选择定义2，当数量较多时，选择定义3作为判别离群点的标准。具体的算法步骤如下：

设学习样本将 X 分为 k 类：$X=\{X_1, X_2, \cdots, X_k\}$，构造样本 X 的覆盖：

(1)覆盖领域的构造

① 求学习样本 X 中样本的最大模 r，并将 X 中的点投影到中心在原点、半径为 r 的球面上；

② 取类别号 $i=1$，构造覆盖 $C(i)$；

③ 若 X_i 中没有尚未覆盖的点，转⑧，否则，任取 X_i 中尚未被覆盖的一点 a_i；

④ 按公式(1)求以 a_i 为中心，阈值 $\theta=d(\omega)$ 的覆盖 $C(a_i)$；

⑤ 求 $C(a_i)$ 所覆盖的点的重心，并将其映射到球面上，设投影点为 a'_i，按步骤④中的公式计算其阈值 θ'，得球形领域 $C(a'_i)$；

⑥ 若 $C(a'_i)$ 覆盖的点数大于 $C(a_i)$ 所覆盖的点数，则令 $a'_i \to a_i$，$\theta' \to \theta$，返回⑤，否则，转⑦；

⑦ 求 ω 的平移点 ω^*，并求对应的球形领域 $C(a_i^*)$，若 $C(a_i^*)$ 覆盖的点数大于 $C(a_i)$ 所覆盖的点数，转⑤，否则，得 $C(i)$ 的一个覆盖，保存覆盖样本数 $n(i)$；若 $i<m$，则 $i+1 \to i$，转③，否则，转⑧；

⑧ 训练结束。

训练结束得到 p 个超球形覆盖领域，即训练样本被分为 k 类不同覆盖集合 $C=\{C_1, C_2, \cdots, C_k\}$

(2)离群点挖掘

①选择阈值 ε 或 ε_1，选择 ε，则转入②，选择 ε_1，则转入③。

②计算 $card(C_j^i)$，获取满足 $card(C_j^i)$ 小于 ε 的所有覆盖领域集合 C'，转入④。

③计算 S_j^i，获取满足 S_j^i 小于 ε_1 的所有覆盖领域集合 C'，转入④。

④输出 C' 领域的覆盖样本，结束。

4 实例分析

我们在 Matlab6.5上设计了利用覆盖算法进行数据挖掘的程序，对来源于 http://www.ics.uci.edu/~mlearn/MLRepository.html 动物库和国旗数据库进行数据进行了离群点挖掘。

4.1 动物数据分析

动物库中共有101种动物，每个动物数据17个属性(类型、有毛发、有羽毛、有卵、有乳、airborne、是水生、是吃肉、有齿、有脊椎、呼吸、有毒、有鳍、有几条腿(legs)、有尾、可圈养、catsize、类别)。各类别动物数量如表1。离群点判别标准选择定义2，ε 取2。

表1

类别	1	2	3	4	5	6	7
数量	41	20	5	13	4	8	10

对所有样本进行覆盖网络学习，网络的覆盖领域如表2。

表2

类别	1	2	3	4	5	6	7
覆盖领域数	1	1	2	1	1	2	3

也就是1、2、4、5类各只有一个覆盖领域，覆盖了全部样本，说明这些类别的动物之间区别很小，无离群点。

第3类5个样本有2个覆盖领域，分别覆盖了4个和1个样本，1个样本是孤点，是 seasnake，与其它4个同类动物相比，只有它有卵、水生、不呼吸。

第6类的2个覆盖领域分别覆盖了2个和6个样本。2个样本是孤点，分别是小昆虫和黄蜂，在这类动物中只有它们会飞，明显区别与其他的6种同类不会飞的动物。

第7类3个覆盖领域分别覆盖了7个、2个和1个样本，1个样本的是蝎子，这一类中只有蝎子无卵且有尾巴，而2个覆盖领域的样本是鼻涕虫和蚯蚓，它们是非食肉动物，而其他同类都是食肉动物。

可见，用覆盖算法获取的离群点十分有效，进一步的分析验证了挖掘出离群点的准确性。

4.2 国旗数据分析

数据库有194个例子(国家或地区)，从提供的信息来看，数据文件包含各个国家以及国旗的具体内容，除去国名，涉及29个属性，包括(洲、地域、面积、人口、语言、宗教、竖条数、横条数、不同颜色数、红、绿、蓝、黄、白、黑、橙、支配颜色、圆圈数、直立十字、X形十字、四分部分数量、太阳或星数量、月牙、三角、图标、文字、左顶部、左底部)，该数据库没有决策属性，是一个典型的信息系统。我们利用覆盖算法进行如下离群点分析：

a)以宗教(Catholic, Other Christian, Muslim, Buddhist, Hindu, Ethnic, Marxist, Others)作为类标记，构造覆盖领域，分析 Muslim 国家国旗库中离群点。

根据覆盖样本分析发现信仰 Muslim 的国家的国旗数据有很大的相似性，亚洲的12个说阿拉伯语的 Muslim 国家分属3个覆盖领域，只有巴林、黎巴嫩是分别被一个覆盖领域所覆盖是离群点(选择 $\varepsilon=1$)，进一步属性分析发现而巴林中人口和面积属性与众不同，其中人口属性值为0，面积也最小，与其他的国家有很大的数量级的差别。另一离群点黎巴嫩的国旗上有图标(Icon)，其属性为1，其余11个国家的该属性全为0。可见，巴林、黎巴嫩数据确有离群之处。

b.选择语言(English, Spanish, French, German, Slavic, Other Indo-European, Chinese, Arabic)作类标记，分析大洋洲说英语国家的国旗数据的离群点。

从覆盖样本得出大洋洲说英语12个国家的国旗数据的离群点国家只有所罗门群岛和巴布亚新几内亚两个国家，覆盖数都为1(选择 $\varepsilon=1$)。进一步比较发现只有所罗门群岛数据

中的 strips 属性为0，其他的11国 strips 的属性为1，并且它是极少数国旗上没有红色且有太阳星的国家。而巴布亚新几内亚国家的信仰独特，信 Ethnic，而其他的11国家信仰都为 Other Christian，国旗属性中的 Green 是0，White 为1，其他国家该属性 Green 为1，White 为0，国旗和信仰同其他国家具有明显得差异。可见挖掘的离群点是可信的。

结论 利用覆盖算法进行离群点挖掘是一种新颖有效的方法，由于构造算法在构造覆盖网络的独特性，使得离群点难以和其它的非离群样本点聚集在同一覆盖领域，只能独自构造覆盖领域，使得应用覆盖领域内的样本分析的方法进行离群点挖掘成为可能。覆盖领域的形成是根据样本自身的特点，没有任何人为的因素，因此根据覆盖算法挖掘出的样本更加真实可靠。同时，覆盖算法适合海量数据和高维数据的处理，因而是一种很好的离群点挖掘的算法。同时覆盖领域的样本本身就是一个很好的模式，对覆盖领域样本的分析可以进行有效的数据挖掘，获取更多的有价值的模式。目前我们正在进行这方面的深入研究。

参 考 文 献

1 Fayyad U, Piatetsky-Shapiro G, Smyth P. Knowledge discovery and data mining: towards a unifying framework. In: Proc. of the 2nd Intl. Conf. on Knowledge Discovery and Data Mining. Portland, Oregon: AAAI Press, 1996. 82～88

2 Barnett V, Lewis T. Outliers in Statistical Data. New York: John Wiley and Sons, Inc., 1994

3 Knorr E M, Ng R T. Algorithms for mining distance-based outliers in large datasets. In: Proc. of the 24th Intl. Conf. on Very Large Data Bases. New York: Morgan Kaufmann, 1998. 392～403

4 Arning A, Agrawal R, Raghavan P. A linear method for deviation detection in large databases. In: Proc. of the 2nd Intl. Conf. on Knowledge Discovery and Data Mining. Portland, Oregon: AAAI Press, 1996. 164～169

5 Zhang Ling, Zhang Bo. A Geometrical Representation of McCulloch-Pitts Neural Model and Its Applications. IEEE Trans. on Neural Networks, 10(4): 925～929

6 张铃，张钹，殷海风. 多层前向网络的交叉覆盖算法. 软件学报，1999，10(7)：737～742

7 Hawkins D. Identification of Outliers. London: Chapman and Hall, 1980

8 吴鸣锐. 大规模模式识别问题的分类器设计研究：[工学博士学位论文]. 北京：清华大学计算机系，2000

9 陶品，张钹，等. 构造型神经网络双交叉覆盖增量学习算法. 软件学报，2003，14(2)：194～201

10 叶少珍，张钹，等. 一种基于神经网络覆盖构造算法的模糊分类器. 软件学报，2003，14(3) 429～434

11 McCulloch W S, Pitts W. A Logical Calculus of the Ideas Immanent in Nervous Activity. Bulletin of Math Biophysics 5. 1943. 115～133

12 郑斌祥，杜秀华，等. 一种时序数据的离群数据挖掘新算法. 控制与决策，2002，17(3)：324～327

（上接第23页）

测井解释；m_4：井下作业。$m' = m_1 \times m_2$，$m'' = m' \times m_3$，$m''' = m'' \times m_4$，全局融合后分析结论的不确定性下降到0，比融合前大大降低，最终的分析结果：强度水淹。系统通过人机交互，给出带可信度的油井开发决策方案。

表1 分布式复合决策融合模型评估结果

多源证据融合	信度函数值(0-1)				评估结果
	$m(u_1)$	$m(u_2)$	$m(u_3)$	$m(\theta)$	
m_1	0.5602	0.3173	0.1035	0.0190	不定
m_2	0.4613	0.4702	0.0401	0.0284	不定
m_3	0.6209	0.2812	0.0713	0.0266	不定
m_4	0.5137	0.3526	0.0513	0.0824	不定
m'	0.6173	0.3644	0.0172	0.0011	不定
m''	0.7779	0.2186	0.0034	0.0001	不定
m'''	0.8291	0.1700	0.0009	0	强水淹

4.3 预测结果的评价

由于模糊综合评判方法的隶属度为某一领域专家主观确定、单一神经网络方法的分析预测精度在偏离训练样本较大时容易产生误差，本文方法与两种方法相比，提高了分析的准确率。每种方法测试100组相同的油井样本，测试结果如表2所示。

表2 测试结果与传统分析方法对照表

评估分类统计	本文方法	模糊综合评判方法	神经网络方法
正确	195	178	183
错误	5	22	17
正确率(%)	97.5	89	91.5

依据剩余油饱和度与水淹级别关系的理论及经验模型，得到两种模型预测的剩余油饱和度对比结果，如图4所示。

从图4中可以看出，本文方法比单一的神经网络方法的预测结果更加符合实际。

结论 将改进的 D-S 证据推理模型、集成神经网络组模型与专家系统融合，提高了分析预测的精度，解决了多源证据矛盾时导致错误结果的问题。

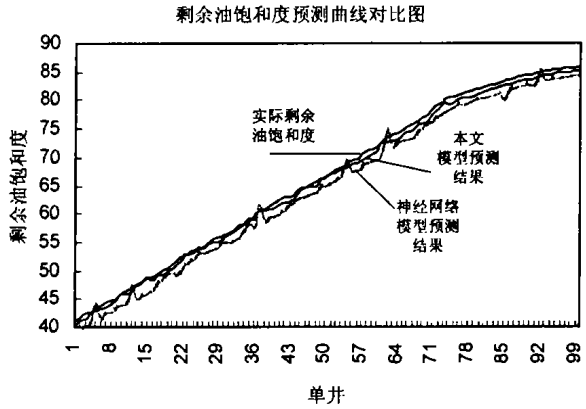


图4 剩余油饱和度预测曲线对比图

将信息融合理论引入到油田剩余油分布及潜力预测的应用研究，为复杂融合系统的工程实现提供了重要的提示。

参 考 文 献

1 Luo R C, Kay M G. Multi-sensor integration and fusion in intelligent systems[J]. IEEE Transaction on Systems, Man and Cybernetics, 1989, 19(5): 901～931

2 Waltz E, Llinas J. Multisensor data fusion. [M]. Boston: Artech House, 1990

3 潘泉，于昕，程咏梅，张洪才. 信息融合理论的基本方法与进展[J]. 自动化学报，2003，29(4)：599～615

4 黎湘，郁文贤，庄钊文，郭桂荣. 决策层信息融合的神经网络模型与算法研究[J]. 电子学报，1997，25(9)：117～120

5 Shafer G. A mathematical theory evidence[M]. Princeton University Press, Princeton, New Jersey, 1976, 26(1)

6 孙怀江，胡忠山，杨靖宇. 基于证据理论的多分类器融合方法研究[J]. 计算机学报，2001，24(3)：231～235

7 Zadeh L A. On the validity of Dempster's rule of combination of evidence[M]. Memo M79/24Univ. of California, Berkeley, 1979

8 Dezert J, Smarandache F. On the generation of hyper-powersets for the DSmT[A]. In: Proc. of the 6th Intl. Conf. on Information Fusion[C]. Cairns, Australia: International Society of Information Fusion (ISIF), 2003: 1118～1125

9 徐凌宇，张德干，赵海. 基于动态相关性挖掘的信息融合方法[J]. 电子学报，2002，30(2)：292～294

10 胡海峰. 基于本体论的领域问题求解方法及在剩余油研究中的应用：[博士学位论文]. 石油大学，北京，2001

11 明仲，蔡树彬，李师贤. 使用约束条件支持领域本体的重用[J]. 计算机科学，2004，31(4)：104～107