

复杂信息系统分布式决策融合模型及应用研究^{*}

魏守智 赵海 王刚 张晓丹

(东北大学信息科学与工程学院 沈阳110004)

摘要 针对复杂信息系统信源为涉及多个不同专业背景的多源异类、异构信息,融合过程必须将多个异构模型集成起来进行联合问题求解的特点,提出了一种新的复杂信息系统分布式层次化决策融合模型,实现了集成神经网络组与专家系统及改进的分布式D-S证据推理模型的融合、特定领域中理论与经验模型与上述模型的融合,改进的D-S证据推理模型解决了当证据矛盾时导致错误结果的问题。油田剩余油分布及潜力预测的应用,表明提高了预测的精度和可靠性,为复杂融合系统的工程实现提供了重要的提示。

关键词 复杂信息系统,信息融合,分布式融合,决策级融合,改进的D-S证据理论

Research of Distributed Decision-Level Fusion Model of Complex Information System with Application

WEI Shou-Zhi ZHAO Hai WANG Gang ZHANG Xiao-Dan

(School of Information Science & Engineering, Northeastern University, Shenyang 110004)

Abstract Aiming at the characteristic of complex information system, whose information source deal with multi-source and different class as well as different structure information in multiple different specialty, whose fusion process must integrate multiple models and carry through combine problem-solving, a new distributed and hierarchy decision-level fusion model of complex information system is proposed. Integrated neural networks groups models fuse expert system and improved D-S evidence reasoning model in distributed and hierarchy system architecture, domain theory models and experiential models fuse above models are implemented. Improved D-S evidence reasoning model solve the problem that evidence combine produce false results when evidences are paradoxical. Oil-field remaining oil distribution and potential forecast system has applied above models, it shows that the precision and reliability of forecast are improved, this paper provides an important guidance to engineer implementation of complex information fusion system.

Keywords Complex information system, Information fusion, Distributed fusion, Decision-level fusion, Improved D-S evidence theory

1 引言

信息融合是现代信息技术与多学科交叉、综合、延拓产生的新的系统科学研究方向。融合是一种形式框架,其过程是用数学方法和技术工具综合不同源信息,目的是得到高品质的有用信息^[1~3]。

决策级融合采用的主要方法有贝叶斯推断、D-S证据理论、模糊综合评判、神经网络等。这些方法各有利弊,其本质是在一定条件下,对具体问题求解,即对信息的处理往往侧重于某一方面的最优^[4]。

针对复杂系统的决策融合问题,如何有机地融合上述方法,成为一个十分重要的研究课题。为此,本文以大庆油田剩余油分布这一复杂问题为研究背景,提出了一种新的分布式层次化决策融合模型,基于本体论设计并实现了大庆油田勘探开发信息融合辅助决策系统。为复杂融合系统的工程实现提供了重要的提示。

2 D-S证据理论及其改进

2.1 D-S证据理论

对于具有主观不确定性判断的多属性评估问题,D-S证据理论是一个融合主观不确定性信息的有效手段^[5,6]。D-S证据推理在多传感器信息融合中的基本应用过程为首先计算各

个证据的基本概率赋值函数 m_i ,信任度函数 Bel_i 和似然函数 Pls_i ,然后用D-S组合规则计算所有证据联合作用下的基本概率赋值函数、信任度函数和似然函数,最后根据一定的决策规则,选择联合作用下支持度最大的假设。

定义1 设 BEL_1 和 BEL_2 是同一识别框架 U 上的两个信任函数, m_1 和 m_2 分别是其对应的基本概率赋值,焦元分别为 $A_1, \dots, A_k$ 和 $B_1, \dots, B_r$,又设

$$K_1 = \sum_{A_i \cap B_j \neq \emptyset} m_1(A_i) m_2(B_j) < 1 \quad (1)$$

则

$$m(C) = \begin{cases} \frac{\sum_{A_i \cap B_j = C} m_1(A_i) m_2(B_j)}{1 - K_1}, & \forall C \subseteq U, C \neq \emptyset; \\ 0, & C = \emptyset \end{cases} \quad (2)$$

在(2)式中若 $K_1 \neq 1$,则 m 确定一个基本概率赋值;若 $K_1 = 1$,则认为 m_1, m_2 矛盾,不能对基本概率赋值进行组合。由定义所给出的证据组合规则称为Dempster组合规则。对于多个证据的组合,可采用该定义对证据进行两两综合。

使用Dempster组合规则要求证据是独立的,当证据相关时,该组合规则就不适用了。D-S理论的一个难题是标准化,它会导致结果与预期的相矛盾,标准化之所以会造成问题,是因为它忽略了所考虑的对象不存在时的信任^[7]。

^{*} 本文得到国家自然科学基金项目(69873007)和大庆油田科技攻关项目的资助。魏守智 博士研究生,研究方向:信息融合理论及应用,智能信息处理;赵海 博士,教授,博士生导师,研究方向:信息融合理论及应用,智能信息处理,计算机网络通信等;王刚 博士,研究方向:信息融合理论及应用,智能信息处理;张晓丹 博士研究生,研究方向:信息融合理论及应用,智能信息处理。

2.2 Dezert-Smarandache (DSmT) 理论

DSmT 理论被认为是 D-S 理论的一般化,当多个证据间存在冲突时,它同样可以可靠地进行处理,其规则组合既可以是不确定的信息,也可以是互相矛盾的信息^[8]。它能够解决传统 D-S 理论解决失败的复杂融合问题。该理论为更广泛类别的融合问题提供了一个特定的框架,因为它与 D-S 理论不同,它的识别框架是完备的。其规则表示如图1所示。

定义2 $\theta = \{\theta_1, \theta_2, \dots, \theta_n\}, 1) \theta_1, \theta_2, \dots, \theta_n \in D^\theta$ 。2) 如果 $A, B \in D^\theta$, 则

$$A \cap B \in D^\theta, A \cup B \in D^\theta$$

$$m(\theta_1) + m(\theta_2) + m(\theta_1 \cup \theta_2) + m(\theta_1 \cap \theta_2) = 1$$

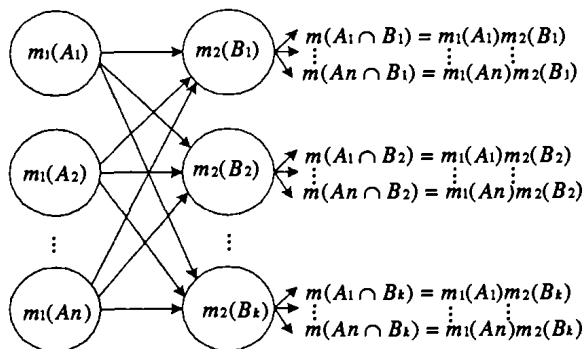


图1 DSmT 规则表示

3 分布式层次化决策融合模型

3.1 神经网络与专家系统的融合

神经网络模型是一种有效的、合理的决策层信息融合模型^[4]。Sherald 认为,专家系统(ES)和人工神经网络(ANN)的协同类似于人的左脑和右脑的协同作用,把二者结合起来可以更好地发挥各自的特长,以解决独自无法解决的问题。ES 与 ANN 的结合可以有多种方式,如 ES 可以用来训练 ANN; ANN 用作 ES 的前端处理机等。本文中的应用系统采用 ES 用作 ANN 的推理解释器的方式。

3.2 神经网络模型与改进的 DSmT 证据推理模型的融合

各个局部融合中心神经网络模型的输出,为各个局部融合中心的分析结论,在全局融合中心运用 DSmT 证据推理模型对这些结论进行组合,可得到更加精确的分析结果。分布式层次化复合决策融合模型的结构如图2所示。

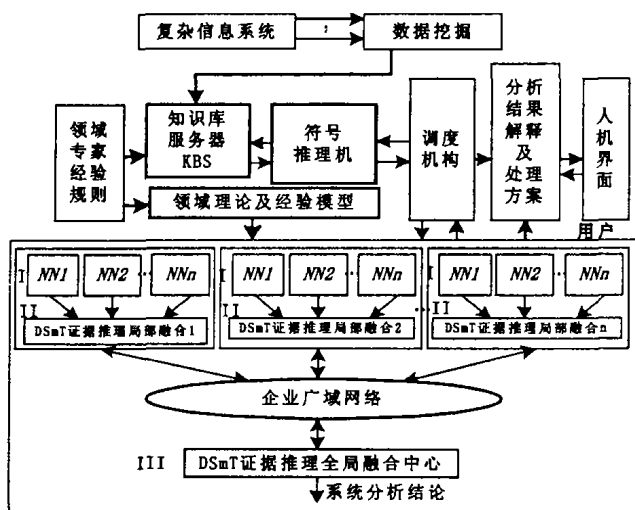


图2 分布式层次化复合决策融合模型结构

图中数据挖掘模块,采用了我们提出的基于动态相关性挖掘的信息融合方法^[9],有效地解决了大规模数据仓库环境下,多因素动态隐含相关性信息的融合问题。

4 大庆油田剩余油分布及潜力预测

4.1 油田勘探开发信息融合辅助决策系统

当今世界石油开采储量的增加3/4以上来自现有储层管理的改善,而仅有1/4弱来自新油田的发现。在油田开发过程中,准确地估算剩余油的饱和度及其分布对于估算一次采油和二次采油的可采储量具有十分重要的意义。目前,国内外描述老油田剩余油的方法主要是运用计算机进行数值模拟,但这一方法仍存在描述周期长、误差大的问题^[10]。大庆油田已经进入高含水后期开发,中好油层部位几乎全部被水淹,个别低渗透差油层含水也较高。为此,我们研究开发了“油田勘探开发信息融合辅助决策系统”,其结构如图3所示。“剩余油分布及潜力预测”是其子系统之一。

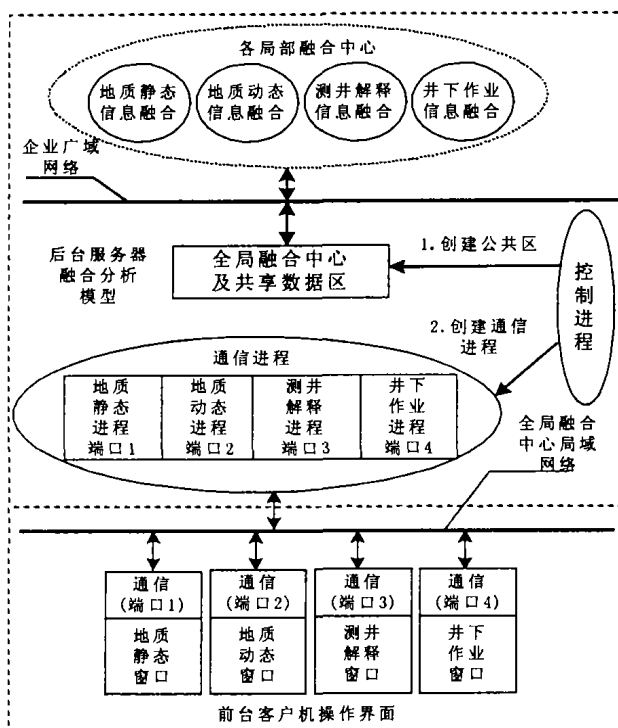


图3 油田勘探开发信息融合辅助决策系统结构

剩余油分布预测涉及地质静态、地质动态、测井解释及井下作业5个专业领域的多种因素,基于油田广域网环境,将系统划分为相应的5个子系统。系统包含很多不确定、不精确因素,故必须采用上述模型中的不确定性推理方法。

基于扩展的 UML 作为支持本体工程的描述语言,我们建立了石油勘探开发信息系统静态本体(知识本体)、动态本体(模型本体)和运行任务本体,实现了不同专业领域间的知识共享、模型集成和重用^[10,11]。

4.2 实际应用与结果分析

我们选择萨尔图油田北部密网试验区100口井进行剩余油分布分析及预测。以某油井为例,四个局部融合中心的 DSmT 证据推理模型对各自集成神经网络组的输出结果进行融合,全局融合中心对四个局部融合中心的局部融合结论进行融合,结果如表1所示。

识别框架为 $U = \{u_1, u_2, u_3\}$, u_1 表示强水淹、 u_2 表示中度水淹、 u_3 表示弱未水淹。表中 m_1 : 地质静态; m_2 : 地质动态; m_3 :

(下转第30页)

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	强度水淹

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。

剩余油饱和度预测曲线对比图

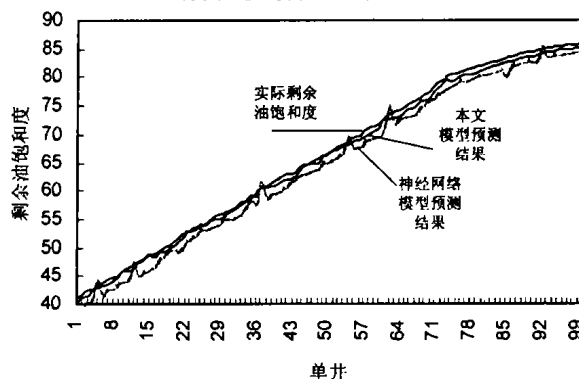


图4 剩余油饱和度预测曲线对比图

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 M79/24 Univ. 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