

基于知识发现创新技术的专家系统新构造

杨炳儒* 宋 威 徐章艳

(北京科技大学信息工程学院, 北京 100083)

摘要 为解决“知识匮乏”这一专家系统中的瓶颈问题, 提出了基于数据库与知识库协同机制的综合过程模型 KD(D&K)及其相关的创新技术, 进而提出了基于知识发现的专家系统 ESKD 新构造. 作为 ESKD 的核心知识获取构件 KD(D&K)由 KDD*和 KDK*两部分组成, 对基于双库协同机制的 KDD*过程模型和基于双基融合机制的 KDK*过程模型分别做了介绍; 给出了 ESKD 的总体框架; 并讨论了 ESKD 的一些子系统和动态知识库系统; 最后, 在农业真实数据库上验证了 ESKD 的有效性和先进性. ESKD 有望把专家系统推向新阶段.

关键词 知识发现 双库协同机制 双基融合机制 KD(D&K) 专家系统 ESKD

自 1965 年 Feigenbaum 研制的 DENDRAL 系统问世以来, 专家系统发展迅速, 已被广泛应用于多个领域. 知识获取是专家系统中的核心组成部分之一. 早期专家系统的知识获取完全依靠领域专家和知识工程师长期共同合作完成. 近来, 一些专家系统已或多或少地具有了自动知识获取的功能, 即系统在运行中不断地从错误和失败中归纳总结经验, 并修改和扩充自身的知识库. 即便如此, “知识匮乏”仍是专家系统中的瓶颈问题^[1,2], 一直制约着专家系统的发展. 知识发现(数据挖掘)是从大量的数据中发现先前未知的、潜在有用的、用户感兴趣的、最终可为用户理解的知识的非平凡过程. 为解决知识获取这一难题, 人们开始通过集成已有的知识发现技术来实现知识的自动获取^[3,4].

我们提出基于知识发现创新技术的专家系统 ESKD 新构造, 其核心思想是把知识发现系统视为认知系统, 研究其自身的潜在规律(机理), 改变知识发现的固有流程, 形成新的知识发现过程模型; 然后再用创新的知识发现机理以及模型作为专家系统新的知识获取构件, 从而丰富和提升了经典专家系统的知识库结构, 形成了全新的动态知识库系统. 新的知识库系统除包含了书本知识、专家经验及通过推理机制得到的知识, 以及通过不断地从错误和失败中进行学习获取的知识外, 更为重要的是为传统的专家系统增加了新的知识获取渠道, 即我们提

收稿日期: 2006-06-02; 接受日期: 2006-08-15

国家自然科学基金重点项目(批准号: 69835001)、国家教育部科技重点项目(批准号: [2000]175)和北京市自然科学基金(批准号: 4022008)资助项目

* 联系人, E-mail: bryang_kd@yahoo.com.cn

出的同时包含数据库中的知识发现(knowledge discovery in database, KDD)和知识库中的知识发现(knowledge discovery in knowledge base, KDK)创新机理和模型的KD(D&K)过程模型, 以及可以在认识发展与时空环境变迁的不同阶段, 使知识不断丰富、升级和深化的技术方法. 这样基于我们提出的知识发现创新技术, 实现了专家系统的新构造. 我们认为, ESKD 的构造对于专家系统的理论研究具有重要意义, 将有可能对新一代专家系统的发展起到重要的推动作用.

本文首先介绍了我们提出的双库协同机制及其派生出的 KDD*新过程模型, 双基融合机制及其派生出的 KDK*新过程模型; 在此基础上, 讨论了将 KDD*与 KDK*有机地融合在一起的新过程模型KD(D&K). 其次, 以KD(D&K)为核心知识获取构件, 提出了基于知识发现的专家系统 ESKD, 给出作为典型示例的诊断型专家系统 ESKD 的总体结构模型及其动态知识库系统. 最后, 通过对比实验及其在农业中的应用, 说明了 ESKD 在解决知识匮乏问题上确实有良好的效果.

1 KD(D&K)

当前国际上知识发现发展的主流是寻求在各类数据库和应用问题的背景下高性能、高扩展性的挖掘算法^[5]. 对于较高层次的框架、乃至理论研究则很少有人问津. 然而知识发现存在一些较难克服的问题, 如海量数据增加引起的算法失效、固有知识库的实时维护、领域专家的局限、先验知识辅助初始发现的聚焦、参数演化问题等. 所有这些KDD研究中所遇到的富有挑战性的问题绝不是单靠算法的研究就能解决的, 它在很大程度上要求助于对知识发现系统自身的内在规律性的认识和揭示. 为解决这些问题: 1) 我们提出双库协同机制, 揭示了数据库与知识库在知识发现过程中, 在各自特定构造下的一一对应关系, 构建了数据库与知识库的内在联系“通道”. 2) 基于KDD与双库协同机制, 我们提出并实现了KDD*过程模型. KDD*有机地沟通与融合了KDD新发现的知识与基础知识库中固有的知识, 同时使领域知识自主地、可具体实现地参与到知识发现过程中. 3) 针对基于数据库的知识发现与基于知识库的知识发现这两个发现过程, 我们提出知识库中的知识发现KDK. 4) 进而又提出了双基融合机制, 其本质在于揭示两个发现过程间的内在联系, 借助KDD来部分地完成KDK的发现任务. 5) 基于KDK与双基融合机制, 我们提出KDK*过程模型. 6) 借鉴协同原理, 提出融合KDD*与KDK*的KD(D&K)过程模型, 将二者统一在一个知识发现系统中, 实现了一种较高的机器智能境界.

1.1 双库协同机制及其诱导的 KDD*过程模型

我们把知识发现系统视为认知系统, 从认知心理学的角度来考察知识发现过程, 提出双库协同机制^[6,7], 主要内容为: 1) 构造两个协调器来模拟认知心理学特征. 用启发型协调器来模拟“创建意向”, 实现系统自主发现知识短缺. 用维护型协调器来模拟“心理信息修复”, 实现知识库实时维护. 2) “定向搜索”和“定向挖掘”, 等效地缩小搜索空间、降低算法的复杂度. 为此, 必须在数据库和知识库的特定构造下, 建立二者之间的某种对应关系.

定理 1 (结构对应定理) 论域 X 的推理范畴 $C_r(N)$ 与完全数据子类结构可达范畴 $C_\infty \langle \gamma, \mathcal{H}(\gamma) \rangle$ 等价(证明略)^[6,7].

该定理给出了知识库中的知识素结点与数据库中“数据子类结构”中的层之间的一一对应关系, 从根本上解决了“定向搜索”与“定向挖掘”的问题.

基于双库协同机制, 提出KDD*新过程模型, 简单地说 $KDD^* = KDD + \text{双库协同机制}$ (其中

“+”表示融合)^[8].

1.2 双基融合机制及其诱导的 KDK*过程模型

如何从知识库中进一步发现更深层次的知识,即知识库中的知识发现KDK,尚无人涉足.针对KDK,我们提出双基融合机制,用数据库与KDD去制约与驱动KDK挖掘过程^[9].

定理 2 (过程模型逻辑等价定理) 设KDK的过程模型为 $M = \langle W, R, M, c \rangle$, KDD的过程模型为 $N = \langle S, F, sup, rel \rangle$. 在依数据子类结构构建数据库和依知识结点网络构建知识库的条件下, M 与 N 各要素间建立了一一对应关系,即 M 与 N 逻辑等价. 其中, W 为知识结点集, R 为知识结点间的认知通达关系(即知识结点间可由归纳推理相连接的关系), M 为正则测度函数, c 为正则确信度函数; S 为数据子类集, F 为数据子类间的可达关系(即数据库中的挖掘途径), sup 为数据子类的支持度, rel 为 F 上的挖掘可信度(证明略)^[9,10].

双基融合机制是由 3 个协调器来具体实现的:

- 1) R 型协调器. 通过综合归纳推理来发现新知识. 知识库的组成包括事实和规则两部分.
- 2) S 型协调器. KDK 发现的规则(特别是难于决断的知识)在进行评价前,先将其送入 KDD*过程中进行定向挖掘,用 KDD*的发掘结果先行评估,若此条规则在 KDD*过程中也可发现,则认为该规则有效的几率较大;反之,则认为此知识缺乏数据支持.
- 3) T 型协调器. 在规则已被 KDK 过程确认后,将产生一个定向搜索进程,搜索知识库中对应位置是否有此生成规则的重复、冗余和矛盾. 这样可以对知识库进行实时维护,做到只对那些最有可能成为新知识的假设进行评价,从而最大限度地减少评价量.

基于双基融合机制,我们提出KDK*模型,简单地讲 $KDK^* = KDK + \text{双基融合机制}$ ^[10].

1.3 KD (D&K)过程模型

以下 3 个问题构成了由上述两个机制共同诱导新过程模型的生长点: 1) 突破基于数据库的知识发现的封闭系统,而与知识库协同起来,由基础知识库制约与驱动 KDD,从而发现新知识; 2) 目前多研究具体挖掘技术,应提升到在宏观背景下多个抽象级和不同知识层面上的知识发现系统的一般性框架的研究; 3) 在“综合基”(数据库和知识库并存)上发现新知识,即将 KDD 与 KDK 有机融合,统一在知识发现的全部运行过程中.

针对上述生长点以及认识与逻辑发展的必然,我们构造了包容 KDD*与 KDK*,而又具特色的新系统 KD(D&K),即具有双库协同机制与双基融合机制的综合型知识发现系统.

简单地讲: $KD(D\&K) = KDD^* + KDK^*$,其总体流程如图 1 所示.

KD(D&K)的基本特征是: 1) 具有多个抽象级和不同知识层面的多层递阶及综合集成的知识发现系统的总体框架与运行机制; 2) 双库(即数据库与知识库)在知识发现过程中具有协同机制; 3) 综合了数据挖掘技术与推理机制(即使得 KDD 与 KDK 统一在知识发现过程中),综合了 KDD 挖掘的知识与知识库的知识,进而再生新知识,特别是在复杂类型数据挖掘方面,比单一 KDD*或 KDK*发现的知识更丰富更全面; 4) 强调知识发现过程中的认知自主性,即突出知识的自动发现(获取),仅在一个发现周期(即一个抽象级)的“始点”与“末端”通过人机交互建立与领域专家的联系.

由此可见, KD (D&K)不是 KDD*与 KDK*的简单叠加,而是区别于 KDD 与 KDK 但又包容两者的、融入双库协同机制与双基融合机制的新过程模型, KD(D&K)与 KDD 及 KDK 特征

上的对比如表 1 所示.

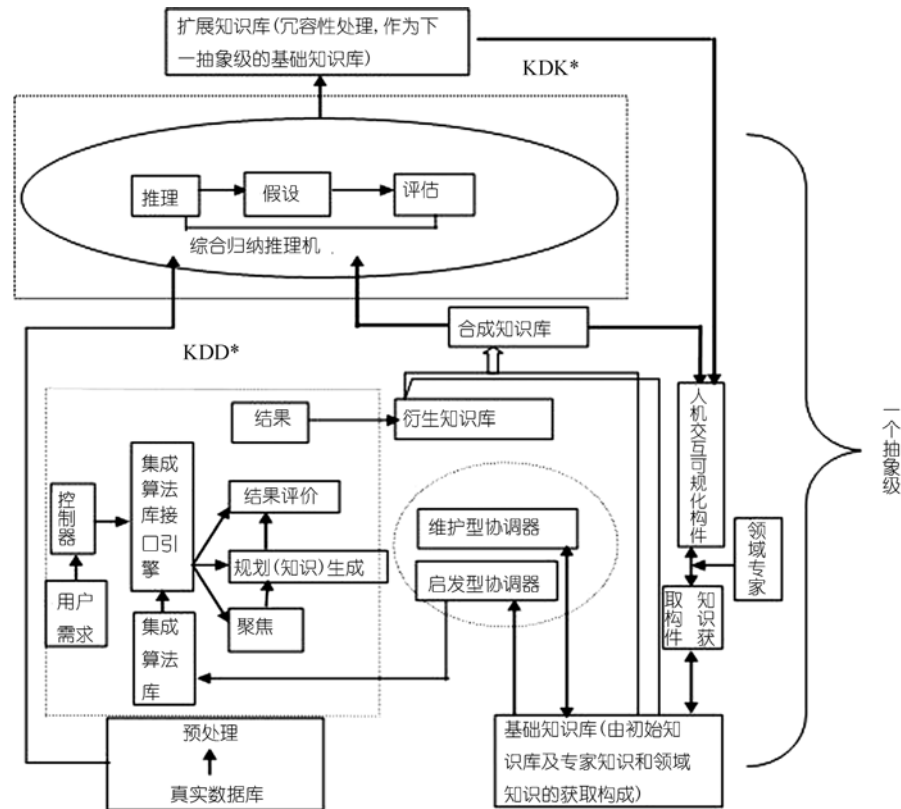


图 1 KD(D&K)总体结构图

表 1 KD(D&K)与 KDD/KDK 的特征对比表

	知识源	KDD 过程	KDK 过程	总体结构	方法、技术	核心概念
KDD 或 KDK	数据库或知识库	基本上形成封闭系统	仅从知识库中利用综合归纳推理机制等发现新知识(规则)	单级、单层 KDD(聚焦—规则生成—结果评价—结果)	KDD 相关的数据挖掘方法与技术	认知自主性
KD (D&K)	数据库、知识库、两库合成,推理机制导出等	通过协调器使知识库制约与驱动 KDD,形成 KDD*,即形成 KDD 的开放系统	在 KDK 中融入双基融合机制,形成 KDK*新结构的新技法	多个抽象级、不同知识层面的多层递阶、综合集成结构	在知识挖掘、表示、评价、优化、冗余性与相容性处理等方法与技术上均有相应的创新	认知自主性+创见意象等(赋予“感兴趣度”、“进化知识”以新的意义)

2 基于知识发现的专家系统 ESKD

专家系统可分为解释型、诊断型、预测型、设计型、规划型等. ESKD 具有较强的通用性,其核心部件 KD(D&K)可根据具体需要与其他部件结合,构成不同类型的专家系统. 本文仅以诊断型的 ESKD 为典型示例进行说明.

2.1 诊断型 ESKD 的总体结构

诊断型 ESKD 的总体结构如图 2 所示.

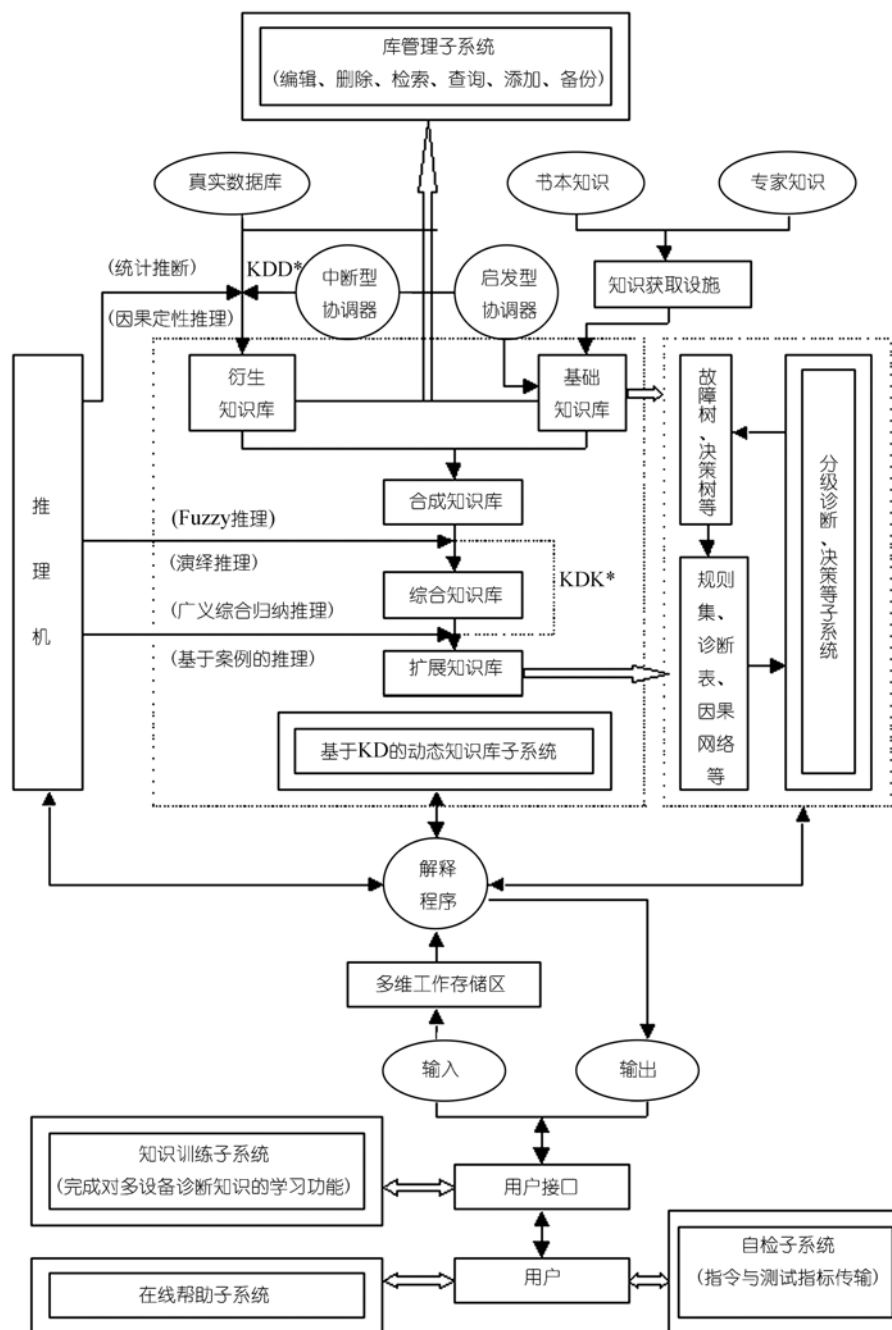


图 2 诊断型 ESKD 总体结构

诊断型 ESKD 总体结构主要包括以下几个模块:

1) 基于知识发现的动态知识库子系统. 这本质上是一个以 $KD(D\&K)$ 为核心的知识发现系统, 作为不同知识层面上知识发现的结果, 使得知识库从原有的由专家经验与书本知识为直接源泉的基础知识库不断产生扩充, 利用在双库协同机制下形成的 KDD^* 和各类推理机制以及 KDK^* , 形成了能处理 Fuzzy 不确定性、随机不确定性及定性信息的具有动态扩展特征的

知识库子系统, 其中以发现关联规则为主, 并用之去修正知识库中原有的故障树、决策树与案例, 以适应复杂的广义诊断问题的求解. 这是本系统的核心, 将在下一小节具体讨论.

2) 知识训练子系统. 既可由专业人员进行经验训练, 又可以通过实例直接得到数据. 从中可以发现知识, 完成专业领域知识的学习功能, 以适应不同用户的需求.

3) 分级诊断、决策等子系统. 先用故障树方法对设备整体进行一系列的指标测试, 以判断其是否有故障. 若有, 则对这些模块逐级进行测试, 找到出现故障的模块后, 即可利用规则库(因果网络, 故障诊断表)方法进入这一模块内部进行测试诊断, 直到找出故障点. 系统通过对设备进行指标测试, 利用正确的相似机制, 依据知识库中的知识, 诊断出设备当前是否正常、故障原因, 并根据决策树提供解决方案.

4) 库管理子系统. 主要是对真实数据库、基础知识库、衍生知识库进行管理, 可进行编辑、删除、检索、查询、添加、备份等管理. 本系统建立良好的界面, 用户可以很方便地实现对各种知识库、数据库的操作管理.

5) 自检子系统. 为避免因测试硬件本身的故障引起错误诊断, 在进行故障诊断之前, 专家系统将对测试硬件进行一次闭环的自我检测.

6) 在线帮助子系统. 为使用户更加有效地使用本系统, 随时能够得到相关信息的帮助.

2.2 诊断型 ESKD 的动态知识库系统

由于把 KD (D&K)作为核心知识获取模块, 解决问题时所用的知识库并非基础知识库, 而是已经过了一系列提升的过程形成的扩展知识库. 其提升过程如下:

1) 基础知识库.

基础知识库中存放了专家经验与书本知识, 基础知识库由规则库 I、故障树 I、决策树 I 与案例库 I 4 个子库构成. 其中规则库包含所有规则; 故障树的目的是判明基本故障, 确定故障的原因、影响和发生概率, 此故障树是用最小割集方法来构建, 可大量减少搜索空间; 决策树用 SLIQ 算法来实现从大型数据库中提取分类规则, 在选择树的分叉时, 同时考虑相关的多个属性, 从而提高产生分类规则的效率; 案例库中存放了以往的典型案例, 可直接利用此库进行子系统诊断时的部分诊断.

2) 衍生知识库.

衍生知识库中存放的是由 KD (D&K) 中的 KDD*发现的新规则.

3) 合成知识库.

合成知识库中存放的是经基础知识库和衍生知识库合成后的知识, 具体方法是: 首先利用衍生知识库中的规则与基础知识库中的规则库 I 进行合成, 形成规则库 II; 然后, 利用规则库 II 去修正基础知识库中的故障树 I、决策树 I 和案例库 I, 形成故障树 II、决策树 II 和案例库 II. 合成知识库的形成过程如图 3 所示.

4) 综合知识库.

综合知识库中存放的是利用模糊推理和演绎推理, 在合成知识库基础上发现的新知识, 同时需对故障树、决策树和案例库进行修正.

5) 扩展知识库.

扩展知识库中存放的是通过广义综合归纳推理机制和基于案例的推理, 在综合知识库的基础上发现的新知识, 即 KD (D&K)中的 KDK*.

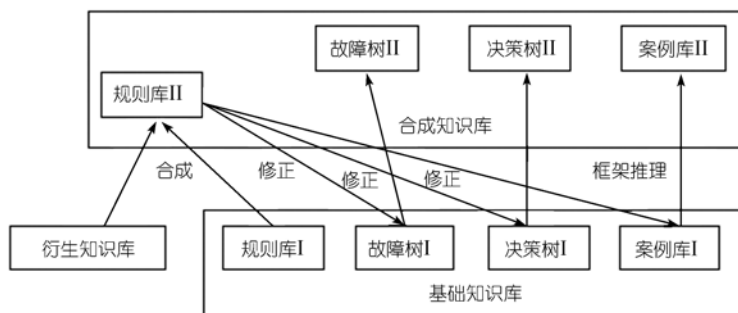


图3 合成知识库的形成过程

以上介绍的知识库经基础—衍生—合成—综合—扩展的提升过程, 只是完成了第 1 个发现(认识)阶段. 如此往复, 在认识发展与时空环境变迁的不同阶段, 不断使知识丰富与升级, 不断使认识深化.

总之, ESKD 用创新的知识发现机理以及模型作为专家系统新的知识获取构件, 从而丰富和提升了经典专家系统的知识库结构, 形成了全新的动态知识库系统. 新的知识库系统既包含了书本知识和专家经验(基础知识库), 也可以通过不断地从错误和失败中进行学习(合成知识库和综合知识库), 通过引入 KD (D&K)过程模型, 为传统的专家系统增加新的知识获取的渠道(衍生知识库和扩展知识库), 而且在认识发展与时空环境变迁的不同阶段, 使知识不断丰富与升级, 认识不断深化.

3 ESKD 与相关工作的对比

目前, 基于知识发现的专家系统可大致分为两类:

其一是利用数据挖掘技术来辅助专家系统的构建, 而不是用于知识获取. 比如, 文献 [11] 利用遗传算法和神经网络对模糊专家系统的关键参数进行优化. 文献 [12] 通过使用软计算(soft computing)的方法来弥补专家系统处理不确定性方面的不足.

另外一类则是将现有的数据挖掘工具直接用于或集成后用于专家系统的知识获取. 比如, 文献 [3] 提出一种面向发电站数据库的、利用 C4.5 等 4 种数据挖掘工具来实现知识获取的专家系统; 文献 [4] 提出一种用决策树来实现知识的自动获取的混合系统, 用于提高表面安装 (SMT) 生产线的产量. 文献 [13] 使用了关联规则和分类算法在历史数据中挖掘知识, 并以此为知识挖掘模块构建了计算机辅助学习专家系统. 通过融合 Bayes 网、朴素 Bayes 等方法, 文献 [14] 提出了一个具有自学习功能的中医药专家系统.

与上面的工作相比, ESKD 的截然不同点与优势在于:

1) ESKD 知识获取构件的理论基础是我们提出的双库协同机制和双基融合机制以及综合型知识发现系统 KD(D&K)过程模型, 而不是对现有的数据挖掘技术的集成和使用. KD (D&K) 过程模型的引入, 为传统的专家系统增加了新的知识获取渠道.

2) ESKD 的动态知识库经历了基础—衍生—合成—综合—扩展的一系列提升过程, 知识存储的数量与质量均极大丰富(传统的知识库系统是 ESKD 动态知识库系统的子系统), 并且其管理系统完备, 能够高智能地发现深层知识与评价知识.

3) 通过协调器的作用和案例学习以及知识训练等方式, 可提高自学习的功能, 不断获得

新的知识加入到动态知识库中；同时基于知识发现的动态知识库及数据库都在时间与空间的延伸中，随着抽象级的增加而再生出适合于环境变迁的新知识，使整个系统具有较强的自适应性。

4) 现有的专家系统往往是专业型的，在通用性方面不尽人意，而 ESKD 则具有较好的通用性，其核心部件 KD (D&K)可根据具体需要与其他部件结合，构成不同类型的专家系统。

5) 专家系统发展至今可大体分为两代：第 1 代专家系统的类型有限，且所处理的问题基本上是可分解的；而第 2 代专家系统不但类型丰富，而且由单一推理机、单一知识库向多推理机、多知识库发展。文献 [15]指出使用混合体系结构是下一代智能系统的研究方向，新的知识获取途径KD (D&K)的使用，丰富了专家系统的结构，使得ESKD有望在下一代智能系统的研究中起到重要作用。

我们在农业真实数据库——庐江螟虫害数据库上进行对比实验。该数据集记录了庐江地区三化螟的数据，共有 1926 条记录，每条记录包括雌虫密度、雄虫密度、最低温度、平均温度、降雨量、日照时数 6 个属性。在实验方法上，我们把 ESKD 与其他两个专家系统进行了对比实验：其一是去掉了 ESKD 中新的知识获取构件 KD(D&K) 模型(记作 ES)；其二是将 ESKD 中的知识获取构件 KD(D&K) 模型换成一般 KDD 模型下的关联规则挖掘构件(记作 ESAR)。在不同的支持度下，比较 3 个系统知识获取构件得到的规则数目。由于专家系统要求得到的知识的可靠程度要高，因此置信度都置为 1(支持度和置信度的设置是针对 ESAR 和 ESKD 而言的，ES 没有使用数据挖掘进行知识获取，不涉及这两个参数)。结果如表 2 和表 3 所示。

表 2 支持度为 1%时，不同系统知识获取构件得到规则数目的比较

规则数目	ES	ESAR	ESKD
后件为“雌虫密度”	19	103	37
后件为“雄虫密度”	18	87	34

表 3 支持度为 30%时，不同系统知识获取构件得到规则数目的比较

规则数目	ES	ESAR	ESKD
后件为“雌虫密度”	19	66	26
后件为“雄虫密度”	18	52	23

由以上结果可知，ESAR 和 ESKD 发现的规则都比 ES 要多，这说明知识发现的引入确实可以在一定程度上解决知识匮乏的问题。对比 ESAR 和 ESKD，我们可以看出 ESAR 得到的规则数目要明显比 ESKD 多，这是由于一般的关联规则挖掘的结果往往存在大量的冗余，严重影响了挖掘结果的可理解性。而 ESKD 则利用了 KD (D&K) 中的“启发型协调器”和“维护型协调器”，较好地实现了定向挖掘和实时维护，从而极大地消除了规则冗余现象。此外，我们发现当支持度较低的时候，ESAR 和 ESKD 都会发现更多的规则。尽管这些在支持度较低时所发现的规则的置信度都为 1，但数据中支持这些规则的记录较少，因此，他们或者是意外规则(支持度较低、可信度较高的规则)，或者是由于噪声引起的，需要领域专家的进一步判断。

4 ESKD 的应用

我们以安徽省合肥农业示范区为基地，将 ESKD 应用于农业施肥、植保领域。

4.1 ESKD 在农业施肥中的应用

在农业生产中，存在着施肥不科学、肥料投入不足、投肥结构及比例失调等肥料施用误区，

严重影响着农作物的优质高产. 构造指导农民科学、合理施用肥料的施肥专家系统是解决施肥问题的一种有效方法. 在施肥专家系统中, 缺乏有效的知识获取工具一直是其应用过程中的瓶颈, 而 ESKD 系统是解决这一瓶颈的有效途径. 图 4 展示了 ESKD 系统在施肥数据上得到的一些结果.

通过在真实施肥数据上使用 ESKD, 我们发现如下一些较有意义的知识: 水稻干物质累积和对硒的累积不同步, 前者高峰在生长中期, 后者以生长后期为主; 另外水稻对硒有一定的生物富集作用. 将这些规则存于知识库, 并根据这些规则, 在农艺措施上应在稻灌浆充实前增施一次硒肥, 而在缺硒和低硒地区施用硒肥, 能显著提高水稻含硒量, 改善其营养品质, 提高其市场竞争力. 这样一方面可以指导人们合理地施加肥料, 另一方面也可以指导肥料生产厂家在不同阶段添加不同的微量元素, 以适应农业生产的需要.

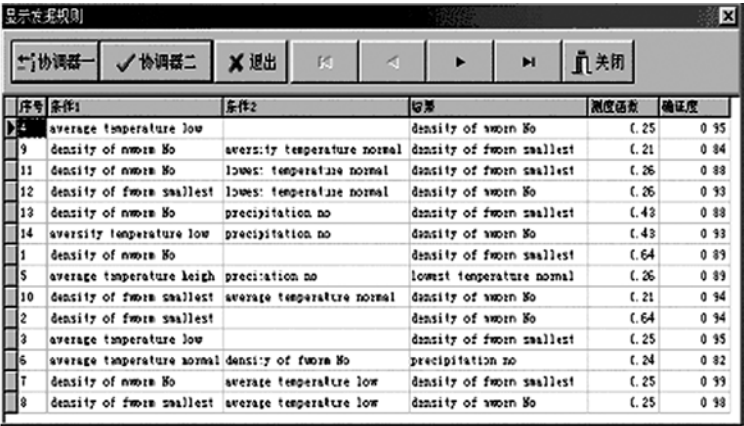


条件一	条件二	条件三	条件四	结论
有机质 (%) >= 3	全氮 (%) >= 0.2	速效钾 (ppm) < 100		肥力分级 = "中"
有机质 (%) >= 3	全氮 (%) >= 0.2	速效钾 (ppm) >= 150		肥力分级 = "高"
有机质 (%) >= 3	全氮 (%) >= 0.2	速效钾 (ppm) [100, 150]	速效磷 (ppm) < 10	肥力分级 = "中"
有机质 (%) >= 3	全氮 (%) >= 0.2	速效钾 (ppm) [100, 150]	速效磷 (ppm) [10, 20]	肥力分级 = "中"
有机质 (%) >= 3	全氮 (%) >= 0.2	速效钾 (ppm) < 100	速效磷 (ppm) >= 10	肥力分级 = "低"

图 4 ESKD 的知识获取构件在施肥数据上得到的部分知识

4.2 ESKD 在农业植保中的应用

对植保数据使用 ESKD 可实现: 1) 发现病虫害诊断知识; 2) 发现气象与病虫害发生规律的关联关系; 3) 预测虫害发生数量; 4) 发现虫害各代发生程度的序贯模式. 在第 3 节中提到的庐江螟虫害数据库上使用 ESKD, 获取到的部分知识如图 5 所示.



序号	条件1	条件2	结论	置信度
1	average temperature low	density of worm No	density of worm No	0.95
9	density of worm No	average temperature normal	density of worm smallest	0.84
11	density of worm No	lowest temperature normal	density of worm smallest	0.88
12	density of worm smallest	lowest temperature normal	density of worm No	0.93
13	density of worm No	precipitation no	density of worm smallest	0.88
14	average temperature low	precipitation no	density of worm No	0.93
1	density of worm No		density of worm smallest	0.89
5	average temperature high	precipitation no	lowest temperature normal	0.89
10	density of worm smallest	average temperature normal	density of worm No	0.94
2	density of worm smallest		density of worm No	0.94
3	average temperature low		density of worm smallest	0.95
6	average temperature normal	density of worm No	precipitation no	0.82
7	density of worm No	average temperature low	density of worm smallest	0.93
8	density of worm smallest	average temperature low	density of worm No	0.98

图 5 ESKD 的知识获取构件在庐江螟虫害数据库上得到的部分知识

我们在庐江螟虫害数据库中发现了一些有价值的规律,例如“降水量很小,雄虫密度很小”、“降水量很小,雌虫密度很小”,所以在降水量大的年份要注意防止三化螟病虫害等(注:这里的“小”是实际数据经语言场、语言值理论处理后的语言值)。

5 结论

为解决“知识匮乏”这一专家系统中的瓶颈问题,在我们提出的基于数据库与知识库协同机制的综合过程模型 KD(D&K)的基础上,构建了基于知识发现的专家系统 ESKD。知识发现创新技术的引入,丰富和提升了经典专家系统的知识库结构,形成了全新的动态知识库系统。新的知识库系统除包含了书本知识、专家经验及通过推理机制得到的知识,以及通过不断地从错误和失败中进行学习获取的知识外,更为重要的是为传统的专家系统增加了新的知识获取渠道,即我们提出的同时包含数据库中的知识发现(KDD)和知识库中的知识发现(KDK)创新机理和模型的 KD (D&K)过程模型,以及可以在认识发展与时空环境变迁的不同阶段,使知识不断丰富、升级和深化的技术方法。ESKD 是一种通用的专家系统新构造,其核心部分 KD (D&K)可以与其他具体的构件相结合形成不同类型的专家系统。通过对比实验以及其在农业中的应用说明了 ESKD 在解决“知识匮乏”问题上确实有良好的效果。我们认为,ESKD 的构造对于专家系统的理论研究具有重要意义,将有可能对新一代专家系统的发展起到重要的推动作用。

参 考 文 献

- 1 Tsumoto S. Automated extraction of medical expert system rules from clinical databases based on rough set theory. *Inf Sci*, 1998, 112 (1-4): 67—84 [\[DOI\]](#)
- 2 杨炳儒. 专家知识的归纳获取. *计算机研究与发展*, 1997, 34 (7): 492—496
- 3 Manuel M L, Guillermo R O. Obtaining expert system rules using data mining tools from a power generation database. *Expert Syst Appl*, 1998, 14(1-2): 37—42 [\[DOI\]](#)
- 4 Sundaresan G, Zhou C, Nelson P C, et al. Hybrid evolutionary algorithm-expert system solution for optimizing SMT placement machines. In: *Proc of IKE'02*. Bogart, GA: CSREA Press, 2002. 674—679
- 5 Mannila H. Theoretical frameworks for data mining. *SIGKDD Expl*, 2000, 1(2): 30—32 [\[DOI\]](#)
- 6 杨炳儒, 王建新. KDD 中双库协同机制的研究 (I). *中国工程科学*, 2002, 4 (4): 41—51
- 7 杨炳儒, 王建新, 孙海洪. KDD 中双库协同机制的研究 (II). *中国工程科学*, 2002, 4 (5): 34—43
- 8 Yang B R. *Knowledge Discovery Based on Inner Mechanism: Construction, Realization and Application*. USA: Elliott & Fitzpatrick Inc, 2004
- 9 杨炳儒, 申江涛, 陈泓婕. 基于知识库的知识发现 KDK 的结构模型与挖掘算法的研究. *中国工程科学*, 2003, 5(6): 49—54
- 10 Yang B R, Shen J T, Song W. KDK based double-basis fusion mechanism and its process model. *Int J Artif Intell Tools*, 2005, 14(3): 399—423 [\[DOI\]](#)
- 11 Perneel C, Themlin J M, Renders J M, et al. Optimization of fuzzy expert systems using genetic algorithms and neural networks. *IEEE Trans Fuzz Syst*, 1995, 3 (3): 300—312 [\[DOI\]](#)
- 12 Baldwin J F, Martin T P, Azvine B. Soft computing for intelligent knowledge-based systems. *BT Tech J*, 1998, 16 (3): 165—179 [\[DOI\]](#)
- 13 Lin Y T, Tseng S S, Tsai C J. The design and implement of a computer-assisted learning expert system. *Int J Comput Proc Oriental Lang*, 2002, 15 (1): 33—61 [\[DOI\]](#)
- 14 Wang X, Qu H, Liu P, et al. A self-learning expert system for diagnosis in traditional Chinese medicine. *Expert Syst Appl*, 2004, 24(4): 557—566 [\[DOI\]](#)
- 15 Brasil L M, Azevedo F M, Barreto J M. Hybrid expert system for decision support in the medical area. *Medic Biol Eng Comput*, 1999, 37 (2): 738—739