

DF 关系数据库的函数依赖研究

刘 芳

(苏州经贸职业技术学院 信息系, 江苏 苏州 215009)

摘 要: 数据间的函数依赖关系一直是关系数据库理论研究中的重要课题, 同样, DF 关系数据库中也存在着函数依赖关系. 首先, 基于 DF 关系理论给出了 DF 关系数据模型的定义, 并介绍了属性值的表示方法; 然后, 用动态模糊等价关系及其隶属函数来描述动态模糊函数依赖, 给出了 DF 数据库中函数依赖的基本定义, 以及基于 DF 蕴涵关系的公理系统, 并证明其推理规则是正确和完备的; 最后讨论了 DF 关系的规范化设计.

关键词: 动态模糊关系; DF 关系数据模型; 动态模糊等价关系; DF 函数依赖; DF 范式

中图分类号: TP311

文献标识码: A

文章编号: 1000-7180(2011)10-0009-05

Research on Functional Dependencies of Dynamic Fuzzy Relational Database

LIU Fang

(Department of Information, Suzhou Institute of Trade & Commerce, Suzhou 215009, China)

Abstract: The functional dependency in data is an important topic in the theory of relational databases. Similarly, the functional dependency also existed in the DF relational database. First, based on dynamic fuzzy relation, dynamic fuzzy relational data model is presented; the representation of dynamic fuzzy data value is introduced. Then, dynamic fuzzy functional dependency is represented by means of dynamic fuzzy equivalent relation and its membership function; the inference rules for dynamic fuzzy functional dependency are described and proved. Finally, the normalization design of dynamic fuzzy relations is discussed.

Key words: dynamic fuzzy relation; dynamic fuzzy relational data model; dynamic fuzzy equivalent relation; dynamic fuzzy functional dependency; dynamic fuzzy normal form

1 引言

在关系数据库理论中引入动态模糊逻辑(Dynamic Fuzzy Logic, DFL)^[1-3]的思想, 运用动态模糊集(Dynamic Fuzzy Sets, DFS)^[2]的表示方法来描述实体, 并存放在数据库中, 由此形成带有动态模糊意义的数据库系统的研究工作已开展多年, 也取得了不少的成果. 文献[4]对 DF 数据模型作了初步的探讨, 文献[5]对 DF 关系数据模型进行了广义的定义, 文献[6]借鉴模糊关系数据操作方法对 DF 关系数据操作进行了初步的描述, 文献[7]讨论了基于 DF 关系数据模型数据库系统的应用. 但上述研究

均没有涉及到 DF 数据间的函数依赖关系的讨论. 而函数依赖是关系数据库和数据建模中所需的关键约束知识, 对数据库建模及实际应用具有重要的意义. 为将函数依赖扩展到 DF 关系数据模型中来, 本文在前述工作的基础上深入探讨了 DF 数据间的函数依赖关系, 给出了 DF 函数依赖的相关概念, 以及 DF 函数依赖的推理规则, 最后讨论了 DF 关系的规范化设计.

2 DF 关系数据模型

2.1 DF 关系及其性质

在动态模糊集中, 如果其论域是几个 DF 集合

的笛卡尔积,则在此论域上的 DF 集称为动态模糊关系 (Dynamic Fuzzy Relation, DFR). 动态模糊关系也是一种 DF 集,但其定义在 n 个论域之上,在实际应用中一般只讨论二元关系^[8].

定义 1 设 $(\tilde{X}, \tilde{X})$ 、 $(\tilde{Y}, \tilde{Y})$ 是两个非空 DF 集,并且 $(\tilde{X}, \tilde{X}) = \{(\tilde{x}_1, \tilde{x}_1), (\tilde{x}_2, \tilde{x}_2), \dots, (\tilde{x}_n, \tilde{x}_n)\}$, $(\tilde{Y}, \tilde{Y}) = \{(\tilde{y}_1, \tilde{y}_1), (\tilde{y}_2, \tilde{y}_2), \dots, (\tilde{y}_m, \tilde{y}_m)\}$, 则称 $(\tilde{X}, \tilde{X}) \times (\tilde{Y}, \tilde{Y})$ 的一个 DF 子集 $(\tilde{R}, \tilde{R})$ 为 $(\tilde{X}, \tilde{X})$ 到 $(\tilde{Y}, \tilde{Y})$ 的一个二元 DF 关系. 对任意的 $((\tilde{x}_i, \tilde{x}_i), (\tilde{y}_j, \tilde{y}_j)) \in (\tilde{X}, \tilde{X}) \times (\tilde{Y}, \tilde{Y}) (i \leq n, j \leq m)$, 其 DF 隶属函数 $(\tilde{R}, \tilde{R})((\tilde{x}_i, \tilde{x}_i), (\tilde{y}_j, \tilde{y}_j)) \in [0, 1] \times [-\infty, +\infty]$ 表示 $(\tilde{x}_i, \tilde{x}_i)$ 和 $(\tilde{y}_j, \tilde{y}_j)$ 之间具有关系 $(\tilde{R}, \tilde{R})$ 的程度.

由定义可知,在 DF 关系中,每一个元素都有一个介于 $[(\tilde{0}, \tilde{0}), (\tilde{1}, \tilde{1})]$ 之间的隶属度,表示该元素隶属于该关系的程度,或称论域元素之间该关系存在的程度. 动态模糊关系具有如下性质:

定义 2 设 $(\tilde{R}, \tilde{R})$ 是定义在 $(\tilde{X}, \tilde{X}) \times (\tilde{Y}, \tilde{Y})$ 上的 DF 关系,若对任意的 $((\tilde{x}, \tilde{x}) \in (\tilde{X}, \tilde{X}), (\tilde{y}, \tilde{y}) \in (\tilde{Y}, \tilde{Y}))$ 都有 $(\tilde{R}, \tilde{R})((\tilde{x}, \tilde{x}), (\tilde{y}, \tilde{y})) = (\tilde{1}, \tilde{1})$, 则称 $(\tilde{R}, \tilde{R})$ 是自反的.

定义 3 设 $(\tilde{R}, \tilde{R})$ 是定义在 $(\tilde{X}, \tilde{X}) \times (\tilde{Y}, \tilde{Y})$ 上的 DF 关系,若对任意的 $((\tilde{x}, \tilde{x}) \in (\tilde{X}, \tilde{X}), (\tilde{y}, \tilde{y}) \in (\tilde{Y}, \tilde{Y}))$ 都有 $(\tilde{R}, \tilde{R})((\tilde{x}, \tilde{x}), (\tilde{y}, \tilde{y})) = (\tilde{R}, \tilde{R})((\tilde{y}, \tilde{y}), (\tilde{x}, \tilde{x}))$, 则称 $(\tilde{R}, \tilde{R})$ 是对称的.

定义 4 设 $(\tilde{R}, \tilde{R})$ 是定义在 $(\tilde{X}, \tilde{X}) \times (\tilde{Y}, \tilde{Y})$ 上的 DF 关系,若 $(\tilde{R}, \tilde{R})$ 的 $(\tilde{\lambda}, \tilde{\lambda})$ 截集 $(\tilde{R}, \tilde{R})_{(\tilde{\lambda}, \tilde{\lambda})} = \{((\tilde{x}, \tilde{x}), (\tilde{y}, \tilde{y})) | (\tilde{R}, \tilde{R})((\tilde{x}, \tilde{x}), (\tilde{y}, \tilde{y})) \geq (\tilde{\lambda}, \tilde{\lambda}), ((\tilde{x}, \tilde{x}), (\tilde{y}, \tilde{y})) \in ((\tilde{X}, \tilde{X}) \times (\tilde{Y}, \tilde{Y}))\}$, 对任意的 $(\tilde{\lambda}, \tilde{\lambda}) \in [(\tilde{0}, \tilde{0}), (\tilde{1}, \tilde{1})]$ 都是传递的关系,则称 $(\tilde{R}, \tilde{R})$ 是传递的.

2.2 DF 关系数据模型的定义

动态模糊关系数据模型 (Dynamic Fuzzy Relational Data Model, DFRDM) 是在上述动态模糊关系的基础上,在普通关系型数据库中引入一个动态模糊隶属函数 (Dynamic Fuzzy Membership Function) 而定义的.

定义 5 设 $D_1, D_2, \dots, D_n$ 为论域 $(\tilde{U}, \tilde{U})$ 上的 n

个 DF 子集,把它们作为所要定义关系的属性值域,则值域 $D_1, D_2, \dots, D_n$ 上的动态模糊关系 $(\tilde{R}, \tilde{R})$ 可定义为一个二元组: $(\tilde{R}, \tilde{R}) = \{(\tilde{r}, \tilde{r}), (\tilde{r}, \tilde{r})(u)\}$, 其中 $(\tilde{r}, \tilde{r})$ 为动态模糊关系 $(\tilde{R}, \tilde{R})$ 的一个实例,其模式为 $(\tilde{R}, \tilde{R})((\tilde{A}_1, \tilde{A}_1), (\tilde{A}_2, \tilde{A}_2), \dots, (\tilde{A}_n, \tilde{A}_n))$, 它是笛卡尔积 $D_1 \times D_2 \times \dots \times D_n$ 的一个子集, $(\tilde{r}, \tilde{r})(u)$ 表示 DF 隶属函数. 对 $(\tilde{r}, \tilde{r})$ 的每个元组 $(\tilde{t}, \tilde{t}) = ((\tilde{u}_1, \tilde{u}_1), (\tilde{u}_2, \tilde{u}_2), \dots, (\tilde{u}_n, \tilde{u}_n)), (\tilde{u}_i, \tilde{u}_i) \in D_i (i = 1, 2, \dots, n)$ 给定一个动态模糊隶属度 $(\tilde{r}(\tilde{t}), \tilde{r}(\tilde{t}))$: $(\tilde{r}(\tilde{t}), \tilde{r}(\tilde{t})) = \min((\tilde{A}_1(\tilde{u}_1), \tilde{A}_1(\tilde{u}_1)), (\tilde{A}_2(\tilde{u}_2), \tilde{A}_2(\tilde{u}_2)), \dots, (\tilde{A}_n(\tilde{u}_n), \tilde{A}_n(\tilde{u}_n)))$, $(\tilde{r}(\tilde{t}), \tilde{r}(\tilde{t})) \in [(\tilde{0}, \tilde{0}), (\tilde{1}, \tilde{1})]$, 表示元组 $(\tilde{t}, \tilde{t})$ 隶属于动态模糊关系实例 $(\tilde{r}, \tilde{r})$ 的程度.

$(\tilde{r}(\tilde{t}), \tilde{r}(\tilde{t}))$ 越接近于 $(\tilde{1}, \tilde{1})$, 表示 $(\tilde{t}, \tilde{t})$ 隶属于 $(\tilde{r}, \tilde{r})$ 的程度越高,若 $(\tilde{r}(\tilde{t}), \tilde{r}(\tilde{t})) = (\tilde{1}, \tilde{1})$, 则表示 $(\tilde{t}, \tilde{t})$ 完全隶属于 $(\tilde{r}, \tilde{r})$; $(\tilde{r}(\tilde{t}), \tilde{r}(\tilde{t}))$ 越接近于 $(\tilde{0}, \tilde{0})$, 表示 $(\tilde{t}, \tilde{t})$ 隶属于 $(\tilde{r}, \tilde{r})$ 程度越低,若 $(\tilde{r}(\tilde{t}), \tilde{r}(\tilde{t})) = (\tilde{0}, \tilde{0})$, 则表示 $(\tilde{t}, \tilde{t})$ 完全不隶属于 $(\tilde{r}, \tilde{r})$.

显然,经典关系数据模型和模糊关系数据模型是 DF 关系数据模型的特例,当 $(\tilde{r}(\tilde{t}), \tilde{r}(\tilde{t})) \in [0, 1]$ 时为模糊关系数据模型;当 $(\tilde{r}(\tilde{t}), \tilde{r}(\tilde{t})) = 1$ 时为经典关系数据模型.

由上述定义可知,DF 关系数据模型表示了两类动态模糊性:属性值的 DF 性和数据间联系的 DF 性. 属性值的 DF 性表示属性值可以是精确的离散值,也可以是论域上的一个 DF 子集;而数据间联系的 DF 性表示元组在关系中的隶属程度,可在关系中用附加列 $(\tilde{r}(\tilde{t}), \tilde{r}(\tilde{t}))$ 表示.

根据不同的应用需要,DF 关系实例上的元组属性值可以采用 DF 数据的各种表示方法,如 DF 隶属函数表示法、DF 区间数表示法和 DF 中心数表示法等^[9]. 当然,对同一属性中数据的表示应当尽可能采用同一种表示方法,使处理更为方便. 表 1 给出了学生课程成绩关系属性值的一个实例.

3 DF 数据间的函数依赖

上面我们简要介绍了 DF 关系的概念及其性质,给出了 DF 关系数据模型的定义,并说明了 DF 数据的表示方法. 接下来,我们就在此基础上探讨 DF 数据间的函数依赖关系. 对于 DF 关系数据模

表 1 学生课程成绩关系属性值的一个实例

学号	姓名	课程号	课程名	成绩	$(\bar{r}(\bar{t}), \bar{r}(\bar{t}))$
01	张山	1001	数学	中等	$\overrightarrow{0.78}$
02	王娟	1001	数学	优秀	$\overrightarrow{0.9}$
04	李明	1003	计算机	$[\overrightarrow{60}, \overleftarrow{69}]$	$\overrightarrow{0.6}$
05	刘红	1004	单片机	$\overleftarrow{45}$	$\overleftarrow{0.45}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

型,数据间的函数依赖也体现了 DF 性,我们将之称为 DF 函数依赖(Dynamic Fuzzy Function Dependencies, DFFD).

3.1 DF 等价关系

DF 函数依赖是基于 DF 等价关系^[10]来定义的,下面先给出 DF 等价关系的定义.

定义 6 论域 $(\bar{U}, \bar{U})$ 上的 DF 关系 $(\bar{R}, \bar{R})$ 称为 $(\bar{U}, \bar{U})$ 上的 DF 等价关系,若对任意的 $(\bar{x}, \bar{x}), (\bar{y}, \bar{y}), (\bar{z}, \bar{z})$ 和 $(\bar{\lambda}, \bar{\lambda}) \in [(0, 0), (\bar{1}, \bar{1})]$ 有:

- (1) 自反的: $(\bar{R}, \bar{R})(\bar{x}, \bar{x}), (\bar{y}, \bar{y}) = (\bar{1}, \bar{1})$
- (2) 对称的: $(\bar{R}, \bar{R})(\bar{x}, \bar{x}), (\bar{y}, \bar{y}) = (\bar{R}, \bar{R})(\bar{y}, \bar{y}), (\bar{x}, \bar{x})$
- (3) 传递的: 若 $(\bar{R}, \bar{R})(\bar{x}, \bar{x}), (\bar{y}, \bar{y}) \geq (\bar{\lambda}, \bar{\lambda}), (\bar{R}, \bar{R})(\bar{y}, \bar{y}), (\bar{z}, \bar{z}) \geq (\bar{\lambda}, \bar{\lambda})$ 则 $(\bar{R}, \bar{R})(\bar{x}, \bar{x}), (\bar{z}, \bar{z}) \geq (\bar{\lambda}, \bar{\lambda})$

定义 7 设 $(\bar{D}, \bar{D}) = \text{dom}((\bar{A}_1, \bar{A}_1)) \times \text{dom}((\bar{A}_2, \bar{A}_2)) \times \cdots \times \text{dom}((\bar{A}_n, \bar{A}_n))$, $(\bar{t}_i, \bar{t}_i), (\bar{t}_j, \bar{t}_j)$ 是 $(\bar{D}, \bar{D})$ 上的两个元组, 即 $(\bar{t}_i, \bar{t}_i), (\bar{t}_j, \bar{t}_j) \in (\bar{D}, \bar{D})$, 对每个属性的值域 $\text{dom}((\bar{A}_i, \bar{A}_i))$, DF 等价关系 $(\bar{R}, \bar{R})$ 的隶属函数记为 $(\bar{R}, \bar{R})(\bar{t}[\bar{A}_i], \bar{t}[\bar{A}_i])$, 则覆盖 $(\bar{D}, \bar{D})$ 的 DF 等价关系 $(\bar{R}, \bar{R})$ 的隶属函数由下式给出:

$$(\bar{R}, \bar{R})(\bar{t}_i, \bar{t}_i), (\bar{t}_j, \bar{t}_j) = \min\{(\bar{R}, \bar{R})(\bar{t}_i[\bar{A}_1], \bar{t}_i[\bar{A}_1]), (\bar{t}_j[\bar{A}_1], \bar{t}_j[\bar{A}_1]), (\bar{R}, \bar{R})(\bar{t}_i[\bar{A}_2], \bar{t}_i[\bar{A}_2]), (\bar{t}_j[\bar{A}_2], \bar{t}_j[\bar{A}_2]), \dots, (\bar{R}, \bar{R})(\bar{t}_i[\bar{A}_n], \bar{t}_i[\bar{A}_n]), (\bar{t}_j[\bar{A}_n], \bar{t}_j[\bar{A}_n])\}$$

3.2 DF 函数依赖及其推理规则

基于上述 DF 等价关系及其隶属函数, 下面定义 DF 函数依赖.

定义 8 设 $(\bar{r}, \bar{r})$ 是论域 $(\bar{U}, \bar{U})$ 上的 DF 关系实

例, $(\bar{X}, \bar{X}), (\bar{Y}, \bar{Y}) \subseteq (\bar{U}, \bar{U})$. 若对任意的 $(\bar{t}_i, \bar{t}_i), (\bar{t}_j, \bar{t}_j) \in (\bar{r}, \bar{r})$ 有 $(\bar{R}, \bar{R})(\bar{t}_i[X], \bar{t}_j[X]) \leq (\bar{R}, \bar{R})(\bar{t}_i[Y], \bar{t}_j[Y])$, 则称 DF 函数依赖(DFFD) $(\bar{X}, \bar{X}) \rightarrow (\bar{Y}, \bar{Y})$ 在 $(\bar{r}, \bar{r})$ 上成立. 其中 $\bar{t}_i[X] = (\bar{t}_i[\bar{X}], \bar{t}_i[\bar{X}]), \bar{t}_j[X] = (\bar{t}_j[\bar{X}], \bar{t}_j[\bar{X}]), \bar{t}_i[Y] = (\bar{t}_i[\bar{Y}], \bar{t}_i[\bar{Y}]), \bar{t}_j[Y] = (\bar{t}_j[\bar{Y}], \bar{t}_j[\bar{Y}])$.

DF 函数依赖的推理规则类似于经典函数依赖的 Armstrong 公理, 现表述如下:

定理 1 令 $(\bar{R}, \bar{R}) = \{(\bar{r}, \bar{r}), (\bar{r}, \bar{r})(u)\}$, 关系模式为 $(\bar{r}, \bar{r})(\bar{A}, \bar{A})$, 且存在 DF 函数依赖集合 $(\bar{F}, \bar{F})$. $(\bar{X}, \bar{X}), (\bar{Y}, \bar{Y}), (\bar{Z}, \bar{Z})$ 是属性集 $(\bar{A}, \bar{A})$ 的子集, 则:

DFFD1(自反公理): 若 $(\bar{Y}, \bar{Y}) \subseteq (\bar{X}, \bar{X})$, 则 $(\bar{X}, \bar{X}) \rightarrow (\bar{Y}, \bar{Y})$.

DFFD2(增广公理): 若 $(\bar{X}, \bar{X}) \rightarrow (\bar{Y}, \bar{Y})$, 则 $(\bar{X}, \bar{X})(\bar{Z}, \bar{Z}) \rightarrow (\bar{Y}, \bar{Y})(\bar{Z}, \bar{Z})$.

DFFD3(传递公理): 若 $(\bar{X}, \bar{X}) \rightarrow (\bar{Y}, \bar{Y}), (\bar{Y}, \bar{Y}) \rightarrow (\bar{Z}, \bar{Z})$, 则 $(\bar{X}, \bar{X}) \rightarrow (\bar{Z}, \bar{Z})$.

DFFD4(并公理): 若 $(\bar{X}, \bar{X}) \rightarrow (\bar{Y}, \bar{Y}), (\bar{X}, \bar{X}) \rightarrow (\bar{Z}, \bar{Z})$, 则 $(\bar{X}, \bar{X}) \rightarrow (\bar{Y}, \bar{Y})(\bar{Z}, \bar{Z})$.

DFFD5(投影公理): 若 $(\bar{X}, \bar{X}) \rightarrow (\bar{Y}, \bar{Y})(\bar{Z}, \bar{Z})$, 则 $(\bar{X}, \bar{X}) \rightarrow (\bar{Y}, \bar{Y}), (\bar{X}, \bar{X}) \rightarrow (\bar{Z}, \bar{Z})$.

DFFD6(伪传递公理): 若 $(\bar{X}, \bar{X}) \rightarrow (\bar{Y}, \bar{Y}), (\bar{Y}, \bar{Y})(\bar{A}, \bar{A}) \rightarrow (\bar{Z}, \bar{Z})$, 则 $(\bar{X}, \bar{X})(\bar{A}, \bar{A}) \rightarrow (\bar{Z}, \bar{Z})$.

引理 1 DF 函数依赖的推理规则(DFFD1 ~ DFFD6)是有效的.

证明 设 $(\bar{r}, \bar{r})$ 为 DF 关系实例, $(\bar{t}_i, \bar{t}_i), (\bar{t}_j, \bar{t}_j)$ 为 $(\bar{r}, \bar{r})$ 中的任意两个元组.

(1) DFFD1: 因为 $(\bar{Y}, \bar{Y})$ 是 $(\bar{X}, \bar{X})$ 的一个子集, 则由定义 7 有 $(\bar{R}, \bar{R})(\bar{t}_i[X], \bar{t}_j[X]) \leq (\bar{R}, \bar{R})(\bar{t}_i[Y], \bar{t}_j[Y])$, 所以 $(\bar{X}, \bar{X}) \rightarrow (\bar{Y}, \bar{Y})$ 成立.

(2) DFFD2: 因为 $(\bar{X}, \bar{X}) \rightarrow (\bar{Y}, \bar{Y})$, 所以 $(\bar{R}, \bar{R})(\bar{t}_i[X], \bar{t}_j[X]) \leq (\bar{R}, \bar{R})(\bar{t}_i[Y], \bar{t}_j[Y])$, 则有 $\min\{(\bar{R}, \bar{R})(\bar{t}_i[X], \bar{t}_j[X]), (\bar{R}, \bar{R})(\bar{t}_i[Z], \bar{t}_j[Z])\} \leq \min\{(\bar{R}, \bar{R})(\bar{t}_i[Y], \bar{t}_j[Y]), (\bar{R}, \bar{R})(\bar{t}_i[Z], \bar{t}_j[Z])\}$, 所以 $(\bar{X}, \bar{X})(\bar{Z}, \bar{Z}) \rightarrow (\bar{Y}, \bar{Y})(\bar{Z}, \bar{Z})$ 成立.

(3)DFFD3: 因为 $(\tilde{X}, \tilde{X}) \rightarrow (\tilde{Y}, \tilde{Y})$ 且 $(\tilde{Y}, \tilde{Y}) \rightarrow (\tilde{Z}, \tilde{Z})$, 所以 $(\tilde{R}, \tilde{R})(t_i[X], t_j[X]) \leq (\tilde{R}, \tilde{R})(t_i[Y], t_j[Y])$ 且 $(\tilde{R}, \tilde{R})(t_i[Y], t_j[Y]) \leq (\tilde{R}, \tilde{R})(t_i[Z], t_j[Z])$, 则有 $(\tilde{R}, \tilde{R})(t_i[X], t_j[X]) \leq (\tilde{R}, \tilde{R})(t_i[Z], t_j[Z])$, 所以 $(\tilde{X}, \tilde{X}) \rightarrow (\tilde{Z}, \tilde{Z})$ 成立.

说明: 上述(1)、(2)、(3)中 $t_i[X] = (\tilde{t}_i[\tilde{X}], \tilde{t}_i[\tilde{X}])$, $t_j[X] = (\tilde{t}_j[\tilde{X}], \tilde{t}_j[\tilde{X}])$, $t_i[Y] = (\tilde{t}_i[\tilde{Y}], \tilde{t}_i[\tilde{Y}])$, $t_j[Y] = (\tilde{t}_j[\tilde{Y}], \tilde{t}_j[\tilde{Y}])$, $t_i[Z] = (\tilde{t}_i[\tilde{Z}], \tilde{t}_i[\tilde{Z}])$, $t_j[Z] = (\tilde{t}_j[\tilde{Z}], \tilde{t}_j[\tilde{Z}])$.

(4)DFFD4: 因为 $(\tilde{X}, \tilde{X}) \rightarrow (\tilde{Y}, \tilde{Y})$, 则由 DFFD2 有 $(\tilde{X}, \tilde{X}) \rightarrow (\tilde{X}, \tilde{X})(\tilde{Y}, \tilde{Y})$; 又因为 $(\tilde{X}, \tilde{X}) \rightarrow (\tilde{Z}, \tilde{Z})$, 则由 DFFD2 有 $(\tilde{X}, \tilde{X})(\tilde{Y}, \tilde{Y}) \rightarrow (\tilde{Y}, \tilde{Y})(\tilde{Z}, \tilde{Z})$, 所以由 DFFD3 有 $(\tilde{X}, \tilde{X}) \rightarrow (\tilde{Y}, \tilde{Y})(\tilde{Z}, \tilde{Z})$ 成立.

(5)DFFD5: 因为由 DFFD1 有 $(\tilde{Y}, \tilde{Y})(\tilde{Z}, \tilde{Z}) \rightarrow (\tilde{Y}, \tilde{Y})$ 和 $(\tilde{Y}, \tilde{Y})(\tilde{Z}, \tilde{Z}) \rightarrow (\tilde{Z}, \tilde{Z})$ 成立, 现又有 $(\tilde{X}, \tilde{X}) \rightarrow (\tilde{Y}, \tilde{Y})(\tilde{Z}, \tilde{Z})$, 所以由 DFFD3 有 $(\tilde{X}, \tilde{X}) \rightarrow (\tilde{Y}, \tilde{Y})$, $(\tilde{X}, \tilde{X}) \rightarrow (\tilde{Z}, \tilde{Z})$ 成立.

(6)DFFD6: 因为 $(\tilde{X}, \tilde{X}) \rightarrow (\tilde{Y}, \tilde{Y})$, 则由 DFFD2 有 $(\tilde{X}, \tilde{X})(\tilde{A}, \tilde{A}) \rightarrow (\tilde{Y}, \tilde{Y})(\tilde{A}, \tilde{A})$ 成立; 又因为 $(\tilde{Y}, \tilde{Y})(\tilde{A}, \tilde{A}) \rightarrow (\tilde{Z}, \tilde{Z})$, 所以由 DFFD3 有 $(\tilde{X}, \tilde{X})(\tilde{A}, \tilde{A}) \rightarrow (\tilde{Z}, \tilde{Z})$ 成立.

引理 2 DF 函数依赖的推理规则(DFFD1 ~ DFFD6)是完备的.

为了证明推理规则的完备性, 下面引入一些基本概念.

定义 9 (DF 函数依赖的逻辑蕴涵) 设 $(\tilde{F}, \tilde{F})$ 是在 $(\tilde{R}, \tilde{R})$ 上成立的 DF 函数依赖集, 若 $(\tilde{X}, \tilde{X}) \rightarrow (\tilde{Y}, \tilde{Y})$ 能由 DFFD 集 $(\tilde{F}, \tilde{F})$ 经推理规则导出, 则称 $(\tilde{F}, \tilde{F})$ 逻辑蕴涵 $(\tilde{X}, \tilde{X}) \rightarrow (\tilde{Y}, \tilde{Y})$.

定义 10 设 $(\tilde{F}, \tilde{F})$ 是在 $(\tilde{R}, \tilde{R})$ 上成立的 DF 函数依赖集, 被 $(\tilde{F}, \tilde{F})$ 逻辑蕴涵的函数依赖全体构成的集合, 称为 DF 函数依赖集 $(\tilde{F}, \tilde{F})$ 的闭包, 记为 $(\tilde{F}, \tilde{F})^+$.

定义 11 设 $(\tilde{A}, \tilde{A})$ 是动态模糊关系实例 $(\tilde{r}, \tilde{r})$ 中所有属性的集合, $(\tilde{X}, \tilde{X}), (\tilde{Y}, \tilde{Y}) \subseteq (\tilde{A}, \tilde{A})$, 属

性集 $(\tilde{X}, \tilde{X})$ 关于 DF 函数依赖集 $(\tilde{F}, \tilde{F})$ 的闭包记为 $(\tilde{X}, \tilde{X})^+$, $(\tilde{X}, \tilde{X})^+ = \{(\tilde{Y}, \tilde{Y}) \mid (\tilde{Y}, \tilde{Y}) \in (\tilde{A}, \tilde{A}) \text{ 且 } (\tilde{X}, \tilde{X}) \rightarrow (\tilde{Y}, \tilde{Y}) \text{ 可由 DF 函数依赖的推理规则导出}\}$.

根据该定义, 可得出 $(\tilde{X}, \tilde{X})^+$ 具有以下 3 个性质:

(1) 若 $(\tilde{X}, \tilde{X}) \rightarrow (\tilde{Y}, \tilde{Y})$ 可由 DF 函数依赖集推出, 则 $(\tilde{Y}, \tilde{Y}) \subseteq (\tilde{X}, \tilde{X})^+$. 因为对每个 $(\tilde{A}_i, \tilde{A}_i) \in (\tilde{Y}, \tilde{Y})$ 都有 $(\tilde{X}, \tilde{X}) \rightarrow (\tilde{A}_i, \tilde{A}_i)$ ($i = 1, 2, \dots, n$) 可由 DF 函数依赖集推出, 即 $(\tilde{A}_i, \tilde{A}_i) \subseteq (\tilde{X}, \tilde{X})^+$, 所以 $(\tilde{Y}, \tilde{Y}) \subseteq (\tilde{X}, \tilde{X})^+$.

(2) $(\tilde{X}, \tilde{X}) \subseteq (\tilde{X}, \tilde{X})^+$. 因为 $(\tilde{X}, \tilde{X}) \subseteq (\tilde{X}, \tilde{X})$, 所以由 DFFD1 有 $(\tilde{X}, \tilde{X}) \rightarrow (\tilde{X}, \tilde{X})$, 从而由(1)得 $(\tilde{X}, \tilde{X}) \subseteq (\tilde{X}, \tilde{X})^+$.

(3) 若 $(\tilde{X}, \tilde{X}) \rightarrow (\tilde{Y}, \tilde{Y})$ 不能从 DF 函数依赖集推出, 则 $(\tilde{Y}, \tilde{Y}) \subseteq (\tilde{X}, \tilde{X})^+$ 不成立. 因为若 $(\tilde{Y}, \tilde{Y}) = \{(\tilde{A}_1, \tilde{A}_1), (\tilde{A}_2, \tilde{A}_2), \dots, (\tilde{A}_n, \tilde{A}_n)\}$, 但所有 $(\tilde{A}_i, \tilde{A}_i)$ 全属于 $(\tilde{X}, \tilde{X})^+$, 则 $(\tilde{X}, \tilde{X}) \rightarrow (\tilde{A}_i, \tilde{A}_i)$ ($i = 1, 2, \dots, n$) 全可从 DF 函数依赖集推出, 所以 $(\tilde{X}, \tilde{X}) \rightarrow (\tilde{A}_1, \tilde{A}_1)(\tilde{A}_2, \tilde{A}_2) \dots (\tilde{A}_n, \tilde{A}_n)$, 即 $(\tilde{X}, \tilde{X}) \rightarrow (\tilde{Y}, \tilde{Y})$ 可从 DF 函数依赖集推出, 此与前提矛盾, 所以 $(\tilde{Y}, \tilde{Y}) \subseteq (\tilde{X}, \tilde{X})^+$ 不成立.

为了证明 DFFD 集所蕴涵的全部依赖都可从 DFFD 集出发用上述推理规则推出, 实际上只要证明不能从 DFFD 集推出的 DF 函数依赖一定不被其所蕴涵即可. 即对于任意给定的不属于 DFFD 集 $(\tilde{F}, \tilde{F})$ 的 DF 函数依赖 $(\tilde{X}, \tilde{X}) \rightarrow (\tilde{Y}, \tilde{Y})$, 如能找到一个 $(\tilde{R}, \tilde{R})$ 上的动态模糊关系实例 $(\tilde{r}, \tilde{r})$, 使 $(\tilde{r}, \tilde{r})$ 满足 $(\tilde{F}, \tilde{F})$ 中的所有 DF 函数依赖, 但不满足给定的 $(\tilde{X}, \tilde{X}) \rightarrow (\tilde{Y}, \tilde{Y})$, 则说明上述推理规则是完备的.

证明 设 $(\tilde{t}_i, \tilde{t}_i), (\tilde{t}_j, \tilde{t}_j)$ 是 $(\tilde{R}, \tilde{R})$ 上的关系实例 $(\tilde{r}, \tilde{r})$ 中的两个元组, 且 $(\tilde{t}_i, \tilde{t}_i)$ 和 $(\tilde{t}_j, \tilde{t}_j)$ 在所有属于 $(\tilde{X}, \tilde{X})^+$ 的属性上取相同的值, 在其他属性上取不同的值.

(1) $(\tilde{r}, \tilde{r})$ 满足的所有 DF 函数依赖在 $(\tilde{F}, \tilde{F})$ 中 设 $(\tilde{W}, \tilde{W}) \rightarrow (\tilde{V}, \tilde{V})$ 是 $(\tilde{F}, \tilde{F})$ 中的一个 DF 函

数依赖. 若 $(\tilde{W}, \tilde{W}) \subseteq (\tilde{X}, \tilde{X})^+$, 则对 $(\tilde{A}_i, \tilde{A}_i) \in ((\tilde{W}, \tilde{W}) - (\tilde{X}, \tilde{X})^+)$, 有 $(\tilde{R}, \tilde{R})((\tilde{t}_i[\tilde{A}_i], \tilde{t}_i[\tilde{A}_i]), (\tilde{t}_j[\tilde{A}_i], \tilde{t}_j[\tilde{A}_i])) = (\tilde{0}, \tilde{0})$, 则 $(\tilde{R}, \tilde{R})((\tilde{t}_i[\tilde{W}], \tilde{t}_i[\tilde{W}]), (\tilde{t}_j[\tilde{W}], \tilde{t}_j[\tilde{W}])) = (\tilde{0}, \tilde{0})$, 由定义 8 知 $(\tilde{W}, \tilde{W}) \rightarrow (\tilde{V}, \tilde{V})$ 成立. 又因为由 $(\tilde{W}, \tilde{W}) \subseteq (\tilde{X}, \tilde{X})^+$ 有 $(\tilde{X}, \tilde{X}) \rightarrow (\tilde{W}, \tilde{W})$ 成立, 则由 DFFD3 得 $(\tilde{X}, \tilde{X}) \rightarrow (\tilde{V}, \tilde{V})$, 所以 $(\tilde{V}, \tilde{V}) \subseteq (\tilde{X}, \tilde{X})^+$. 即 $(\tilde{W}, \tilde{W}), (\tilde{V}, \tilde{V}) \subseteq (\tilde{X}, \tilde{X})^+$, 又由前提 $(\tilde{R}, \tilde{R})((\tilde{t}_i[\tilde{W}], \tilde{t}_i[\tilde{W}]), (\tilde{t}_j[\tilde{W}], \tilde{t}_j[\tilde{W}])) = (\tilde{R}, \tilde{R})((\tilde{t}_i[\tilde{V}], \tilde{t}_i[\tilde{V}]), (\tilde{t}_j[\tilde{V}], \tilde{t}_j[\tilde{V}]))$, 所以 $(\tilde{r}, \tilde{r})$ 满足 $(\tilde{W}, \tilde{W}) \rightarrow (\tilde{V}, \tilde{V})$.

(2) $(\tilde{r}, \tilde{r})$ 不满足 $(\tilde{X}, \tilde{X}) \rightarrow (\tilde{Y}, \tilde{Y})$

由 $(\tilde{r}, \tilde{r})$ 的定义有 $(\tilde{t}_i[\tilde{X}], \tilde{t}_i[\tilde{X}]) = (\tilde{t}_j[\tilde{X}], \tilde{t}_j[\tilde{X}])$, 若 $(\tilde{r}, \tilde{r})$ 满足 $(\tilde{X}, \tilde{X}) \rightarrow (\tilde{Y}, \tilde{Y})$, 必有 $(\tilde{t}_i[\tilde{Y}], \tilde{t}_i[\tilde{Y}]) = (\tilde{t}_j[\tilde{Y}], \tilde{t}_j[\tilde{Y}])$, 由 $(\tilde{t}_i, \tilde{t}_i)$ 和 $(\tilde{t}_j, \tilde{t}_j)$ 的定义, $(\tilde{Y}, \tilde{Y}) \subseteq (\tilde{X}, \tilde{X})^+$, 则可用推理规则由 $(\tilde{F}, \tilde{F})$ 中推导出 $(\tilde{X}, \tilde{X}) \rightarrow (\tilde{Y}, \tilde{Y})$, 此与前提矛盾. 所以 $(\tilde{r}, \tilde{r})$ 不满足 $(\tilde{X}, \tilde{X}) \rightarrow (\tilde{Y}, \tilde{Y})$.

由 DF 蕴涵的定义知, 只要有一个适合 DFFD 集的关系 $(\tilde{r}, \tilde{r})$ 不适合 $(\tilde{X}, \tilde{X}) \rightarrow (\tilde{Y}, \tilde{Y})$, 就可知 $(\tilde{X}, \tilde{X}) \rightarrow (\tilde{Y}, \tilde{Y})$ 不被 DFFD 集所蕴涵. 所以, 完备性得证.

3.3 DF 关系的规范化

类似于经典关系数据库范式的概念, 也可以将 DF 关系数据库模式分为不同的 DF 范式.

设 $(\tilde{R}, \tilde{R}) = \{(\tilde{r}, \tilde{r}), (\tilde{r}, \tilde{r})(u)\}$, 关系模式为 $(\tilde{r}, \tilde{r})(\tilde{A}, \tilde{A})$, $(\tilde{A}, \tilde{A})$ 为属性集, 现将不同的 DF 范式定义如下:

定义 12 如果 $(\tilde{r}, \tilde{r})(\tilde{A}, \tilde{A})$ 中每个属性不可分解, 属性值为 DF 子集或精确的离散值, 则称该 DF 关系模式属于第一 DF 范式, 记作 $(\tilde{r}, \tilde{r})(\tilde{A}, \tilde{A}) \in 1\text{DFNF}$, 而 $(\tilde{r}, \tilde{r})$ 是满足第一 DF 范式的 DF 关系.

定义 13 如果 $(\tilde{r}, \tilde{r})(\tilde{A}, \tilde{A}) \in 1\text{DFNF}$, 且 $(\tilde{r}, \tilde{r})(\tilde{A}, \tilde{A})$ 中每个非主属性完全动态模糊函数依赖于模式的键, 则称该 DF 关系模式属于第二 DF 范式, 记作 $(\tilde{r}, \tilde{r})(\tilde{A}, \tilde{A}) \in 2\text{DFNF}$, 而 $(\tilde{r}, \tilde{r})$ 是满足第

二 DF 范式的 DF 关系.

定义 14 设 $(\tilde{F}, \tilde{F})$ 是关系模式 $(\tilde{r}, \tilde{r})(\tilde{A}, \tilde{A})$ 的 DF 函数依赖集, $(\tilde{X}, \tilde{X}) \cup (\tilde{Y}, \tilde{Y}) \subseteq (\tilde{F}, \tilde{F})$, 如果对 $(\tilde{F}, \tilde{F})$ 中每个非平凡的依赖 $(\tilde{X}, \tilde{X}) \rightarrow (\tilde{Y}, \tilde{Y})$, 都有 $(\tilde{X}, \tilde{X})$ 是 $(\tilde{r}, \tilde{r})(\tilde{A}, \tilde{A})$ 的超键或 $(\tilde{Y}, \tilde{Y})$ 是主属性, 则称该 DF 关系模式属于第三 DF 范式, 记作 $(\tilde{r}, \tilde{r}) \in 3\text{DFNF}$, 而 $(\tilde{r}, \tilde{r})$ 是满足第三 DF 范式的 DF 关系.

限于篇幅, 这里只给出了动态模糊第一范式、第二范式和第三范式的基本概念, 关于它们的求解算法类似于经典关系数据库中相应范式的求解算法, 这里就不再展开讨论了.

4 结束语

正如函数依赖是传统关系数据库的基础, DF 函数依赖也是 DF 关系数据库的基础. 研究 DF 关系数据库的函数依赖对 DF 数据的规范化、查询优化以及 DF 数据间的关联关系等的研究都有重要的理论和实际意义. 本文基于 DF 关系理论, 针对 DF 关系数据模型, 提出了 DF 函数依赖的概念, 研究了 DF 函数依赖的蕴涵问题, 并给出了 DF 函数依赖的推理规则, 从而扩展了关系数据库中函数依赖的概念.

参考文献:

- [1] 李凡长, 朱维华. 动态模糊逻辑及其应用[M]. 昆明: 云南科技出版社, 1997: 11-16.
- [2] Li Fanzhang. Dynamic fuzzy sets and its applications [M]. Kunming: Yunnan Science Press, 1997: 4-10.
- [3] 李凡长, 刘贵全, 余玉梅. 动态模糊逻辑引论[M]. 昆明: 云南科技出版社, 2005: 8-14/87-90.
- [4] Filev D. System approach to dynamic fuzzy models[J]. International Journal of General Systems, 1992, 21(3): 311-337.
- [5] 李凡长. 动态模糊关系数据模型研究[J]. 小型微型计算机系统, 2002, 23(9): 1107-1109.
- [6] 石鲁生. 面向动态模糊数据的关系数据库设计方法及应用研究[D]. 苏州: 苏州大学, 2007.
- [7] 刘芳, 李凡长. 动态模糊关系数据模型及其应用研究[C]//模糊逻辑与计算智能研究进展, 2005 年中国模糊逻辑与计算智能联合学术会议论文集. 合肥: 中国科学技术大学出版社, 2005: 135-139.
- [8] 刘芳. DF 关联规则及其挖掘算法的研究[J]. 微电子学与计算机, 2009, 26(10): 93-98.

(下转第 17 页)

测系统的目的是最大化正确率的同时减少误报率. 在本实验中,正确率为 96.2%,误报率为 2.9%.

表 1 最优单层前馈全连通网络的测试结果

		实际类型		
		正常	欺诈 A	欺诈 B
输出	正常	9 661	12	1
	欺诈 A	8	117	0
	欺诈 B	2	0	213

4.3 双层全连通前馈网络的实验

一般来说,多层前馈网络比单层前馈网络效果好. 进化方法也可以优化多层前馈网络的结构.

图 5 展示了双层前馈网络的进化过程,其中所有的网络都是全连通的. 在本实验中,种族数量 M 为 20,最大代数 G_{max} 为 400. 实验独立运行了 15 次. 显然在进化过程中,隐藏层结点数量和连接数量有着相似的特征.

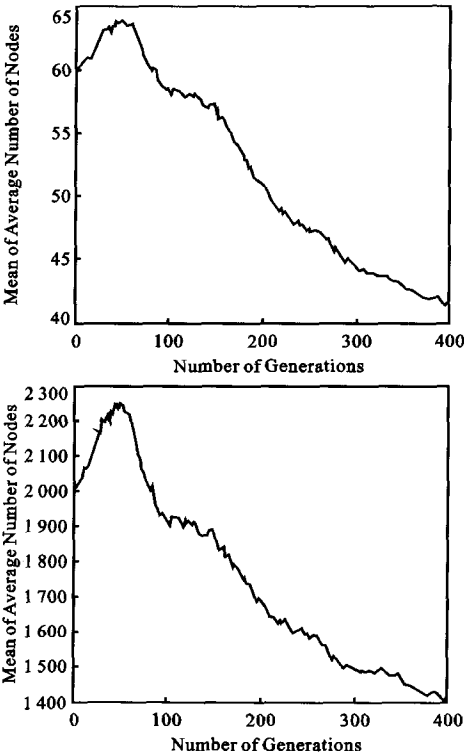


图 5 双层前馈网络的结构和准确度

我们在最终的种族中选择了最佳的个体,并在

测试集上进行测试,结果如表 2 所示. 正确率为 97.4%,误报率为 2.9%.

表 2 最优双层前馈全连通网络的测试结果

		实际类型		
		正常	欺诈 A	欺诈 B
输出	正常	9 661	9	1
	欺诈 A	8	120	0
	欺诈 B	2	0	214

5 结束语

近年来,信用卡欺诈越来越猖獗. 为了保障发卡机构和持有人的利益,本文提出了一种基于进化神经网络的信用卡欺诈检测技术,并展示了在特定的欺诈检测问题中,神经网络结构是如何通过进化算法得到自动化的优化选择. 然而,网络结构的进化过程需要较长的计算时间,因此如何通过更好的编码策略以及基因操作来优化进化效率需要进一步的研究.

参考文献:

[1] Odri S V, Petrovacki D P, Krstonosic G A. Evolutional development of a multilevel neural network[J]. Neural Networks, 1993,6(4):583-595.

[2] Yao X, Liu Y. A new evolutionary system for evolving artificial neural networks[J]. IEEE. Tran. Neural Networks, 1997(8):694-713.

[3] Barse E L, Kvarnstrom H, Jonsson E. Combining fraud and intrusion detection—meeting new requirements[C]// Proc. Nordic Workshop. Secure IT systems. Ireland: Cork, 2000.

[4] Barse E L, Kvarnstrom H, Jonsson E. A synthetic fraud data generation methodology[C]// Lecture Notes in Computer Science, ICICS 2002, Laboratories for Information Technology, Singapore, Springer Verlag, 2002.

[5] Margare H Dunham. Data mining introductory and advanced topics[M]. Boston: Prentice Hall, 2003.

[6] Yao X. A review of evolutionary artificial neural networks[J]. Int. J. Intell. Syst., 1993,8(4):539-567.

作者简介:

凌晨添 男,(1987—),硕士. 研究方向为信息检索与处理、机器学习.

(上接第 13 页)

[9] 李凡长,郑家亮. 基于 DFS 的专家数据库系统设计[J]. 计算机工程, 2001,27(3): 96-99.

[10] 刘芳,李凡长. DF 关系及其在数据聚类中的应用研究[J]. 计算机工程与应用,2009,45(26): 138-146.

作者简介:

刘 芳 女,(1965—),硕士,副教授. 研究方向为数据库技术、动态模糊系统、智能软件.