

基于 BP 神经网络的属性匹配方法研究 *

强保华^{1,2} 陈凌¹ 余建桥¹ 吴开贵² 吴中福²

(西南大学信息学院 重庆 400716)¹ (重庆大学计算机学院 重庆 400044)²

摘要 为了实现异构数据库的数据共享,关键的问题就是要找出数据库间的相同属性。目前主要采用的方法是通过比较所有的属性来实现属性的相似性匹配,但是当同一属性用不同数据类型表示时,由于描述属性的元数据信息和取值信息的极大差异性,这些方法就不能找出相同的属性。并且将不同数据类型描述的属性放在一起匹配,还会造成属性数据之间的干扰,影响匹配结果的准确性。为此,本文提出一种基于 BP 神经网络的二步检查法属性匹配算法。该算法中属性首先根据数据类型进行分类,然后用分类后的属性集分别多次训练神经网络,并对每次的匹配结果求交集作为最终的属性匹配结果,进行两阶段检查,即二步检查法。该算法能有效地消除不一致信息的干扰,降低神经网络的规模,并且可以实现不同数据类型的属性集之间属性匹配过程的并行计算。实验结果显示本文提出的方法能明显地提高系统的运行效率、属性匹配的查准率和查全率。

关键词 BP 神经网络,属性匹配,二步检查法,异构数据库

Research on Attribute Matching Approach Based on Neural Network

QIANG Bao-Hua^{1,2} CHEN Ling¹ YU Jian-Qiao¹ WU Kai-Gui² WU Zhong-Fu²

(College of Information, Southwest Agricultural University, Chongqing 400716)¹

(Department of Computer Science and Engineering, Chongqing University, Chongqing 400044)²

Abstract In order to realize data sharing, identifying corresponding attributes is an important issue in heterogeneous databases. The main approaches at present use the characteristics describing attributes to evaluate the similarity of attributes by comparing all attributes. But these approaches can not present correct results due to the obvious difference of metadata and value information describing attributes when the same attribute is expressed using different data types, and result in incorrect attributes matching for the interference among attributes with different data types. So two phase checking algorithm based on BP neural network is presented to realize attributes matching, in which attributes are required to be categorized according to data types, and the BP neural networks are trained several times respectively using the categorized attributes, and the final attributes matching results are the intersection of every time matching results. This algorithm can resolve the interference among attributes with different data types, and decrease the size of BP neural network, and realize the parallel computation of attributes matching. The experimental results show our approach can improve the system performance, the precision ratio and recall ratio of attributes matching obviously.

Keywords BP neural network, Attributes matching, Two phase checking algorithm, Heterogeneous databases

1 引言

随着信息技术的发展,越来越多的应用需要访问异构数据库中的数据。为了实现异构数据库中的数据共享和互操作,首要的问题就是找出异构数据库中的相同属性。文[1~8]讨论了如何找出数据库中相同属性的问题,主要是通过描述属性的元数据信息和取值信息来进行属性的匹配,这些方法都是很有用的,因为属性的元数据信息和取值信息很好地体现了属性的语义。但这些方法是通过比较所有的属性来实现属性间的相似性匹配,特别当现实世界中同一属性用不同数据类型来表示时,由于描述属性的元数据信息和取值信息的极大差异性,就不能找出不相同(似)的数据类型(后面给出属性的数据类型之间相同(似)性的定义)描述的另一属性。例如,

关系模式: *Student(Sno, Sname, Sage, Sdept)*

假如描述属性 Sno 的元数据及值信息的数据指标为:

(数据类型,长度,是否为主键,取值的约束条件,平均值,

最小值,最大值)

当描述属性 Sno 的数据类型为整型时,各数据指标的取值为:

(整型,4,主键,非空,95030,95001,95059) (1)

当描述属性 Sno 的数据类型是字符串型时,最大值、最小值及平均值是指实际的字符串长度的取值。各数据指标的取值为:

(字符串型,5,主键,非空,5,5,5) (2)

由于不相同(似)数据类型描述同一属性时数据指标取值的巨大差异性,比较式(1)和(2)不能够找出相同的属性,并且将不同数据类型描述的属性放在一起匹配,存在不是相同属性但属性的数据指标的取值可能相同(似)而得到匹配的情形,从而使属性匹配的准确率和系统的运行效率降低。

为此,本文提出一种根据属性的数据类型分类训练的神经网络方法来实现异构数据环境下属性的匹配。该方法明显优于文[4,5]的神经网络方法,为了减小不同数据类型描述的属性之间匹配的干扰,提高匹配结果的准确率,首先对属性根

* 基金资助项目:国家自然科学基金项目(70371030);重庆市教委基金项目(040212)。强保华 博士研究生,讲师,主要研究方向为异构数据库集成、信息集成、网络安全。陈凌 硕士研究生,主要研究方向为分布式数据库、网络安全。余建桥 博士,教授,主要研究方向为人工智能、分布式数据库。吴开贵 博士后,主要研究方向为信息安全。吴中福 教授,博士生导师,主要研究方向为计算机与网络安全。

据数据类型进行了分类,并给出了用不同的初始权值和阈值多次训练神经网络并对匹配结果求交集的二步检查法。实验结果显示本文提出的方法能显著提高系统的运行效率、属性匹配的查准率(precision)和查全率(recall)。

2 基于 BP 神经网络的属性匹配方法

有很多进行异构数据环境下相同属性的确定方法是通过固定的规则来实现的,如文[9~11]。规则匹配的方法并未考虑属性数据自身的特点,需要先验知识来确定描述属性的元数据和值的信息的权重。但属性的元数据及值的信息的权重很难准确量化,也不能找到一个通用的固定的属性匹配规则。神经网络则不同,它是通过具体的实例进行训练而不是根据确定的规则进行编程,神经网络方法不需要先验知识,能够充分考虑到数据的自身特点且有较强的推广能力。实验结果证明本文给出的基于 BP 神经网络的二步检查法属性匹配算法比文[4,5,9]更有效。

本文给出的基于 BP 神经网络的属性匹配方法是在属性根据数据类型分类的基础上进行的。

2.1 根据数据类型对属性分类

根据描述属性的数据类型的异同,将数据类型分为三大类:数值型(包含 bigint, int, smallint, tinyint, decimal, numeric, float, real);字符串型(包含 char, varchar, text, nchar, nvarchar, ntext);稀有型(包含 money, smallmoney, datetime, smalldatetime, timestamp)。由于照片数据牵涉到图像识别与检索问题,所以本文不考虑内容为图像的属性的匹配问题。

定义(相同(似)数据类型) 在数据类型的三大类划分中每一类内部的数据类型之间,称为相同(似)的数据类型。

根据数据类型对属性进行分类后,神经网络的规模随之减小,训练也更加容易,这样将花费较少的时间进行属性的匹配,能明显地提高系统的性能。关于该问题的并行计算可行性,我们已在文[12,13]中进行了详细的阐述。

2.2 建立描述属性的数据指标体系

为了提高属性匹配的准确率,描述属性的数据指标体系中应包含属性的元数据信息、值的信息及其它语义信息。从文[4]对描述属性的各个数据指标的敏感度分析结果看,起作用比较大的指标主要有以下指标:

(字符类型、数值类型、稀有类型、数据类型长度、是否允许为空、精度、小数位数、最小值、最大值、平均值、差异系数、标准差、数字字符比率)

以上数据指标作为神经网络的输入。当描述属性的数据类型是字符串型时,最小值、最大值、平均值、差异系数、标准差分别指实际的字符串长度的值。对于日期型我们通过提取年份的函数来计算年份的最小值、最大值、平均值、差异系数及标准差。

作为神经网络的输入信息,各数据指标的值若不是数值型表示必须首先量化为数值表示,再通过转换函数将这些信息转换到区间[0,1]之间,转换函数采用 $f(\text{length}) = 2 * (1 / (1 + k^{-\text{length}})) - 0.5$,其中 $k = 1.01$ 。

2.3 基于 BP 神经网络的二步检查法属性匹配算法

由于神经网络是由输入通过连接权及阈值的计算得出网络的输出,理论上存在着网络的输入不同却有着相同(似)的输出情形。特别当属性较多时,输出向量的维度增大,神经网络会给出很多不正确的匹配结果,如果直接根据这些输出结果判断,其结果不能令人接受。文[4,5]解决了用神经网络进行异构数据库间的属性匹配问题,但只是根据一次训练的

神经网络给出匹配结果,其中有很多结果是不正确的。

为了解决该问题,本文提出一种基于 BP 神经网络的二步检查法属性匹配算法。算法的基本思想是:首先对属性根据数据类型分类,然后在各分类的属性集内部对神经网络单独训练二至三次,训练结束后,求出待匹配属性分别在对应的神经网络上的输出,最后对输出结果求交集,交集即为最终的属性匹配结果。

算法:基于 BP 神经网络的二步检查法属性匹配算法

假设有两个待进行属性匹配的数据库,分别用 DB1 和 DB2 表示。令字符串类型的属性集合为 C,数值类型的属性集合为 N,稀有类型的属性集合为 R。

步骤 1 将 DB1 中的属性根据数据类型按 2.1 节给出的分类规则分为 C1、N1 和 R1,将 DB2 中的属性分为 C2、N2 和 R2,并按 2.2 节介绍的方法对数据进行规范化。

步骤 2 选取训练数据集,假定这里选用 DB1 中的 C1、N1 及 R1 作为训练数据集。根据 C1、N1 及 R1 的特点,分别建立对应的神经网络 BPN1、BPN2 和 BPN3。

步骤 3 分别用 C1、N1 及 R1 训练对应的神经网络 BPN1、BPN2 和 BPN3,各进行 3 次训练。

```
For(i=1; i<=3; i++)
```

```
{
    随机生成网络的权重和偏置,用 C1 训练 BPN1,记下训练结束后网络的各项参数;
    随机生成网络的权重和偏置,用 N1 训练 BPN2,记下训练结束后网络的各项参数;
    随机生成网络的权重和偏置,用 R1 训练 BPN3,记下训练结束后网络的各项参数;
};
```

步骤 4 分别用 DB2 中的测试数据集 C2、N2 及 R2 在对应的已训练结束后的神经网络 BPN1、BPN2 和 BPN3 上进行属性是否相似的匹配,各进行 3 次匹配。

```
For(i=1; i<=3; i++)
```

```
{
    求出 C2 在神经网络 BPN1 上的输出集合 S1i;
    求出 N2 在神经网络 BPN2 上的输出集合 S2i;
    求出 R2 在神经网络 BPN3 上的输出集合 S3i;
};
```

步骤 5 求出结果集。预先置 S1₀、S2₀、S3₀ 为全集。

```
For(i=1; i<=3; i++)
```

```
{
    S1i = S1i ∩ S1i-1;
    S2i = S2i ∩ S2i-1;
    S3i = S3i ∩ S3i-1;
};
```

S1₃、S2₃、S3₃ 即为 DB1 和 DB2 中相似的属性。本文提出的基于 BP 神经网络的二步检查法属性匹配算法基于以下事实:真正一致的信息,不论一次或多次训练神经网络,输出结果始终是稳定的,但不一致的信息在不同的连接权和阈值的 BP 网络中输出结果是不一致的(每次训练神经网络的连接权和阈值都是随机生成的)。通过二步检查法,能够去除匹配结果中不相关数据的干扰,使匹配结果的准确性显著提高,在第 3 部分的实验中将得到很好的验证。

3 实验结果

该部分,对本文提出的基于 BP 神经网络的二步检查法属性匹配算法进行验证。实验环境为:CPU 为 PIV 2.6G,内存 512M,操作系统为 Windows XP,数据库为 SQL Server 和 ACCESS,生成神经网络的函数采用 net=newff(minmax(p), [11,10], {'tansig','logsig'}, 'traingdx'),训练参数为 net.trainParam.show=50; net.trainParam.lr=0.01; net.trainParam.epochs=20000; net.trainParam.goal=3.0e⁻⁶。

用于训练神经网络的属性数据见表 1(来源于 SQL Serv-

er 2000),用于与表 1 中属性进行匹配的属性数据见表 2 (来源于 ACCESS 2000),所有属性数据已按 2.2 节的方法进行了规范化。通过对比分析表 Employees 与表 Orders 中的数据,表中共有三对相似属性。为了确定表 1 与表 2 中的相似属性,先用表 1 中的属性训练神经网络,然后用表 2 中的属性进行匹配。首先,我们按照文[4, 5]中给出的传统的方法训练神经网络并进行属性匹配,即不分类的情形。该种情形下 BP 神经网络的平均训练时间为 25.146 秒,平均训练周期为 5591 次。作为对比,将属性根据数据类型分类,然后分别训练神经网络三次,则分类后每次神经网络的平均训练时间为 8.688 秒(包含各分类的神经网络训练时间之和),平均训练周期为 1554。分类后神经网络的训练时间和训练周期明显减少。

衡量本文给出的方法的优劣,仅从训练时间和训练周期来评价是不够的,还必须对匹配结果的查准率和查全率进行评价,表 3 给出了神经网络的具体匹配结果,整个表从分类与不分类两大项进行评价。首先,不分类情形下三次训练的神经网络的匹配结果位于前三行,第二步检查法的匹配结果是前

三次神经网络的匹配结果的交集。其次,分类情形下的属性匹配分别在数值型属性、字符型属性和稀有型属性训练的神经网络内部进行,第二步检查法的匹配结果仍是前三次神经网络的匹配结果的交集,具体匹配结果的查准率、查全率及 BP 神经网络训练时间对比见表 4。从表 3 和表 4 可以看出,第二步检查法的查准率明显高于传统的一次训练的神经网络方法,分类的情形又明显高于不分类的情形。从表 4 中可以看出,如果对分类情形下第二步检查法花费的时间与文[4,5]中不分类情形下的一般方法花费的时间进行对比,花费的时间基本持平,但匹配结果的准确率大大提高,查全率都能达到 100%,查准率从一般方法的 27.4%达到了本文给出的第二步检查法的 75%。

从实验结果可以看出,本文提出的算法找出了最理想结果。我们对本文提出的方法在 SQL Server 2000 和 ACCESS 2000 数据库上进一步做了验证,结论是一致的。相信随着属性的增多,在大型数据库之间处理属性匹配的问题,进而解决数据库集成,基于 BP 神经网络的第二步法属性匹配算法会有更大的优势。

表 1 Employees 表中数据规范化后的的属性信息

属性	数据类型	输入向量	输出向量
EmployeeID	数值型	(0, 0.0992, 0, 0.0198, 0, 0.0497, 0, 0.0050, 0.0447, 0.0249, 0.0027, 0.0136, 1)	(1, 0, 0, 0, 0, 0, 0, 0, 0)
LastName	字符型	(0.4600, 0, 0, 0.0990, 0, 0, 0, 0.0199, 0.0447, 0.0354, 0.0011, 0.0080, 0)	(0, 1, 0, 0, 0, 0, 0, 0, 0)
FirstName	字符型	(0.4600, 0, 0, 0.0497, 0, 0, 0, 0.0199, 0.0398, 0.0287, 0.0010, 0.0060, 0)	(0, 0, 1, 0, 0, 0, 0, 0, 0)
Title	字符型	(0.4600, 0, 0, 0.1482, 1, 0, 0, 0.0646, 0.1190, 0.0981, 0.0007, 0.0142, 0)	(0, 0, 0, 1, 0, 0, 0, 0, 0)
TitleofCourtesy	字符型	(0.4600, 0, 0, 0.1237, 1, 0, 0, 0.0149, 0.0199, 0.0155, 0.0005, 0.0017, 0)	(0, 0, 0, 0, 1, 0, 0, 0, 0)
Address	字符型	(0.4600, 0, 0, 0.2899, 1, 0, 0, 0.0745, 0.1480, 0.1050, 0.0012, 0.0251, 0.1257)	(0, 0, 0, 0, 0, 1, 0, 0, 0)
City	字符型	(0.4600, 0, 0, 0.0745, 1, 0, 0, 0.0298, 0.0398, 0.0326, 0.0006, 0.0036, 0)	(0, 0, 0, 0, 0, 0, 1, 0, 0)
Country	字符型	(0.4600, 0, 0, 0.0745, 1, 0, 0, 0.0010, 0.0149, 0.0127, 0.0010, 0.0026, 0)	(0, 0, 0, 0, 0, 0, 0, 1, 0)
BirthDate	稀有型	(0, 0, 0.9630, 0.0398, 1, 0, 0, 1, 1, 0, 0.0451, 0.7780)	(0, 0, 0, 0, 0, 0, 0, 0, 1, 0)
HireDate	稀有型	(0, 0, 0.9630, 0.0398, 1, 0, 0, 1, 1, 0, 0.0043, 0.7780)	(0, 0, 0, 0, 0, 0, 0, 0, 0, 1)

表 2 Orders 表中数据规范化后的的属性信息

属性	数据类型	输入向量
OrderID	数值型	(0, 0.0992, 0, 0.0199, 0, 0.0497, 0, 1, 1, 0.0001, 0.8310, 1)
EmployeeID	数值型	(0, 0.0992, 0, 0.0199, 1, 0.0497, 0, 0.0050, 0.0448, 0.0199, 0.0031, 0.0120, 1)
CustomerID	字符型	(0.4600, 0, 0, 0.0249, 1, 0, 0, 0.0249, 0.0249, 0.0249, 0, 0, 0)
ShipName	字符型	(0.4600, 0, 0, 0.1960, 1, 0, 0, 0.0398, 0.1680, 0.0878, 0.0014, 0.0248, 0)
ShipCity	字符型	(0.4600, 0, 0, 0.0745, 1, 0, 0, 0.0199, 0.0745, 0.0402, 0.0018, 0.0148, 0)
ShipPostalCode	字符型	(0.4600, 0, 0, 0.0497, 1, 0, 0, 0, 0.0447, 0.0273, 0.0016, 0.0088, 0.9700)
OrderDate	稀有型	(0, 0, 0.9630, 0.0398, 1, 0, 0, 1, 1, 0, 0.0035, 0.7780)
Freight	稀有型	(0, 0, 0.6330, 0.0398, 1, 0.0942, 0.0199, 0.0010, 0.9999, 0.3707, 0.0074, 0.5234, 1)

表 3 Orders 表中属性与 Employees 表中属性的匹配结果 (相似度阈值为 0.9)

属性数据类型		数值型		字符型			稀有型		
属性		OrderID	EmployeeID	CustomerID	ShipName	ShipCity	ShipPostalCode	OrderDate	Freight
不分类	第 1 次训练的神经网络的匹配结果	Title, Address, BirthDate	EmployeeID	Firstname, City	Title	Country	null	HireDate	Title, Address, BirthDate
	第 2 次训练的神经网络的匹配结果	LastName, Address, BirthDate	EmployeeID	City	Title	null	null	HireDate	LastName, Address, BirthDate
	第 3 次训练的神经网络的匹配结果	Firstname, Title, City, BirthDate	EmployeeID	City	Title	Country	null	HireDate	Firstname, BirthDate
	第二步法匹配结果	BirthDate	EmployeeID	City	Title	null	null	HireDate	BirthDate
分类	第 1 次训练的神经网络的匹配结果	EmployeeID	EmployeeID	Firstname, City	Title	null	Firstname, Address, City	HireDate	BirthDate
	第 2 次训练的神经网络的匹配结果	null	EmployeeID	City	Title	Country	TitleofCourtesy, Country	HireDate	BirthDate
	第 3 次训练的神经网络的匹配结果	EmployeeID	EmployeeID	Firstname, TitleofCourtesy, City	LastName	LastName, Address, Country	Address, City	HireDate	BirthDate
	第二步法匹配结果	null	EmployeeID	City	null	null	null	HireDate	BirthDate

FCD15444-1.htm

- 3 王相海. 基于小波的图像和视频可分级编码研究: [南京大学博士后研究工作报告]. 2001
- 4 王相海. 基于小波的图像可分级编码研究. 小型微型计算机系统, 2003, 24(9): 1677~1681
- 5 W3C. Portable Network Graphics (PNG) Specification (Second Edition) Information technology - Computer graphics and image processing - Portable Network Graphics (PNG); Functional specification. ISO/IEC 15948; 2003 (E). W3C Recommendation 10 November 2003. <http://www.w3.org/TR/2003/REC-PNG-20031110>
- 6 ISO/IEC, ISO/IEC 14495-1; 1999; Information technology - Lossless and near-lossless compression of continuous-tone still images: Baseline, December 1999
- 7 ISO/IEC, ISO/IEC 14496-2; 1999; Information technology - Coding of audio-visual objects - Part 2; Visual, December, 1999
- 8 PNG. <http://zh.wikipedia.org/wiki/PNG>
- 9 Taubman D. High performance scalable image compression with

- EBCOT. IEEE Trans Image Proc, 2000, 9(7): 1158~1170
- 10 刘方敏, 吴永辉, 于建新. JPEG2000 图像压缩过程及原理概述. 计算机辅助设计与图形学学报, 2002, 14(10): 905~912
- 11 Marcellin M W, Bilgin A. JPEG2000; Highly Scalable Image Compression. In: Proceedings of 2001 International Conference on Information Technology; Coding and Computing (ITCC2001), Las Vegas, Nevada, 2001. 268~272
- 12 王相海, 张福炎. 静态图像编码研究进展. 计算机研究与发展, 2001, 38(11): 1315~1326
- 13 Shapiro J M. Embedded image coding using zerotrees of wavelet coefficients. IEEE Transactions on Signal Processing, 1993, 41(12): 3445~3462
- 14 Fukunaga S, et al. MPEG-4 video verification model version 16.0. WSO/IEC document JTC1/SC29/WG11 N3312, March 2000/ Noordwijkerhout
- 15 Santa-Cruz D, Ebrahimi T. An analytical study of JPEG 2000. In: Proc. of the Intl. Conf. on Image Processing (ICIP), Vancouver, Canada, Sept. 2000. 10~13

(上接第 251 页)

表 4 查准率、查全率及 BP 网络训练时间对比 (训练精度为 $MSE=3.0 \times 10^{-6}$)

		平均查准率 (precision)	平均查全率 (recall)	平均训练时间 (秒)
一般方法	不分类的情形	27.4%	100%	25.146
	分类的情形	30.2%	100%	8.688
二步检查法	不分类的情形	50%	100%	75.438
	分类的情形	75%	100%	26.064

结论 本文给出了一种在异构数据环境下进行属性匹配的方法, 该方法能够提高异构数据库间属性匹配的准确率, 明显高于文[4, 5, 9]中属性匹配的查准率和查全率, 减少了人为判断的负担。相信随着属性的增多, 查准率会进一步得到提高。基于 BP 神经网络的二步检查法属性匹配算法有如下优点:

①属性根据数据类型分类后, 描述属性的数据指标趋于一致, 向量的维度减少, 这样进行神经网络训练和匹配将花费更少的时间, 特别是在进行大型数据库集成时是有积极意义的;

②属性根据数据类型分类后, 由于相同属性的确定是在相同(似)的数据类型描述的属性内部进行, 所以整个属性匹配过程可以在不相同(似)的数据类型之间进行并行计算, 能够明显地提高系统效率;

③属性根据数据类型归类后, 可以消除不同数据类型描述的属性之间的匹配干扰, 提高属性匹配结果的准确率;

④二步检查法通过多次训练神经网络, 并对每次的匹配结果求交集, 可以去除匹配结果中不相关数据的干扰, 避免了人为对匹配结果的重新评价, 减轻了用户的负担, 大大提高了匹配结果的准确率。

我们下一步的工作是在大型异构数据库环境下进行实验并且解决不相同(似)的数据类型描述的同属性的匹配问题, 实现异构数据库间属性匹配及集成的高效率。

参 考 文 献

- 1 Hayne S, Ram Su. Multi-user view integration system(MUVIS): An expert system for view integration. In: Proc. in the 6th Intl. Conf. on Data Engineering, 1990. 402~409
- 2 Salton G, Yang C S, Yu C T. A theory of term importance in au-

- tomatic text analysis. Journal of the American Society for Information Science, 1975, 26(1): 33~44
- 3 Benkley S S, Fandozzi J F, Housman E M, et al. Data element tool-based analysis (DELTA): [Technical Report MTR 95B0000147]. The MITRE Corporation, Bedford, MA, 1995
- 4 Li W-S, Clifton C, Liu S-Y. Database integration using neural networks; implementation and experiences. Knowledge and Information Systems, Springer-Verlag London Ltd, 2000, 2: 73~96
- 5 Li W-S, Clifton C. Semantic integration in heterogeneous databases using neural networks. In: Proc. of the 20th VLDB Conf. Santiago, Chile, 1994
- 6 Premerlani W J, Blaha M R. An approach for reverse engineering of relational databases. Communications of the ACM, 1994, 37(5): 42~49
- 7 Yu C, Sun W, Dao S, Keirse D. Determine relationships among attributes for interoperability of multi-database systems. In: Proc. of RIDE-IMS. IEEE, March, 1991
- 8 Sheth A, Larson J. Federated database systems for managing distributed heterogeneous, and autonomous databases. Computer Surveys, 1990, 22(3): 183~236
- 9 Li W-S, Clifton C. Using field specifications to determine attribute equivalence in heterogeneous databases. In: Third Intl. Workshop on Research Issues on Data Engineering; Interoperability in Multidatabase Systems, IEEE, Vienna, Austria, 1993. 174~177
- 10 Larson A, Navathe S B, Elmasri R. A theory of attribute equivalence in database with application to schema integration. Trans. on Software Engineering, 1989, 15(4): 449~463
- 11 Sheth A, Larson J, Cornelio A, et al. A tool for integrating conceptual schemas and user views. In: Proc. of the 4th Intl. Conf. on Data Engineering, Los Angeles, CA, IEEE, 1988
- 12 强保华, 吴中福, 陈凌, 吴开贵, 余建桥. 异构数据库环境下语义集成过程的并行计算方法研究. 计算机科学, 2004(9)
- 13 强保华, 吴中福, 余建桥, 陈凌, 吴开贵. 一种基于属性信息熵的实体匹配方法研究. 计算机工程, 已录用