

文章编号: 2095-2163(2021)09-0022-06

中图分类号: TP391

文献标志码: A

数据空间中一种基于步进的后映射(SPM)迭代方法

吉露¹, 曹斌²

(1 贵州大学 大数据与信息工程学院, 贵阳 550025; 2 中铝智能科技发展有限公司, 杭州 310000)

摘要: 数据空间是一个自我维持和自我管理的系统, 需要支持中介模式及其模式映射之间的不确定性。通过将概率和可靠度分配给模式映射, 有些映射的结果并不完全正确, 这将不利于用户的精确查询请求。本文提出了一种 SPM(step post-mapping) 映射方法, 利用可能性理论对可靠映射进行管理, 将可靠映射集划分为可能性分布的子集, 并给每个子集分配一个递归的可能性函数。使用铝电解真实数据集, 通过查准率、查全率和 F 值评估算法, 并将其与现有的方法进行对比分析, 验证该方法的可行性。

关键词: 数据空间; 可靠度; 模式映射

An iterative method of step-by-step post-mapping (SPM) in data space

JI Lu¹, CAO Bin²

(1 College of Big Data and Information Engineering, Guizhou University, Guiyang 550025, China;

2 Chinalco Intelligent Technology Development Co. Ltd., Hangzhou 310000, China)

[Abstract] The data space is a self-sustaining and self-managing system. The uncertainty between mediation patterns and their schema mappings needs to be supported. By assigning probability and reliability to pattern mappings, some of the mappings are not exactly correct. This is harmful for the user's accurate query requests. It is necessary to increase the probability and reliability of the mapping. In this paper, a SPM(Step Post-Mapping) mapping method is proposed to manage the reliable mapping using possibility theory. The set of reliable mappings is divided into subsets of probability distribution, and a recursive probability function is assigned to each subset. The actual dataset of aluminum electrolysis is used to compare and analyze the existing methods through the accuracy, recall and F value to verify the feasibility of the method.

[Key words] dataspace; degree of reliability; schema mapping

0 引言

当前, 数据正呈现着海量、多样和动态的特点, 使得数据集成和数据管理需要遵循 pay-as-you-go 的模式^[1], 数据的管理是在管理过程中逐步的完善, 这有别于传统的数据库系统以及数据集成系统中高成本、高功能的集成和管理技术。因此, 数据空间以“pay-as-you-go”的模式来管理这些海量、动态、异构数据为目标的新型数据管理方式应运而生^[2-3]。然而, 由于数据间的模式异构带来数据的检索效率下降, 一定程度上阻碍了数据空间的发展。

将数据空间看作是一个虚拟的空间, 在其中管理许多的数据源, 而不去考虑其结构和位置^[4]; 这样做的目的是不需要任何领域专家协助的情况下, 提供基本的功能服务, 信息的检索和关键字检索等。然后创建自动的语义映射, 这将导致多种可能的语义映射, 映射的结果有些是正确的, 有些并不完全正确, 这带

来了第一类不确定映射, 概率模式映射和可靠中介模式映射用来处理多个可能映射之间的不确定^[5]。在概率映射和可靠中介模式映射中, 系统需要选择一个阈值, 数据空间中语义映射的数量较大, 选择一个阈值成为一项不确定的任务, 会带来数据丢失。

针对模式匹配, COMA 系统通过组合多个匹配器来提高两个输入模式之间的语义映射^[6]; 另一个系统 SF 以两个图模式作为输入, 并在模式的对应节点之间生成映射作为输出^[7]。PORSCHE 从一组基于树的多个输入模式中提供了一个中介模式^[8]。这些系统都只是接受两个源模式作为输入, 而其它系统接受多个模式但通常都带有领域专家的反馈, 数据空间的应用领域中, 用户可能没有足够的技能来操作映射器或提供精确的映射, 并且连接到数据空间的源模式具有异构的结构, 可以是文件, 关系数据库, XML 存储库, web 页面等。因此, 数据空间需要一种新的数据集成方法。数据空间中的集成是

作者简介: 吉露(1996-), 女, 硕士研究生, 主要研究方向: 信息系统、数据空间技术; 曹斌(1963-), 男, 博士, 研究员, 硕士生导师, 主要研究方向: 工业控制、信息安全和智能管理。

通讯作者: 曹斌 Email: cb2027@126.com

收稿日期: 2021-07-04

“pay-as-you-go” 现收现付方式^[9]。对于数据空间这样的数据管理场景下, 黄毅芳等人将本体的概念引入到数据空间中^[10]; 在引入本体概念的情况下, 苏畅提出一种基于 hash 映射, 将记录映射为空间中的一个点, 进行映射结果的分块^[11]; 寇月等人提出了语义项映射方法, 考虑数据源间的语义关联^[12]; Kuicheu 等提出了一种将概率映射与可靠度 (RMedMap) 结合的方法, 通过设定阈值来确定可靠的映射集, 容易带来数据的丢失^[13]。本文提出了一种后映射方法 (SPM) 来处理这种映射的不确定性, 不需要选择阈值。引入可能性理论, 将可靠的映射集合进行排序并划分为可能性分布的子集, 给每个子集分配一个递归可能度函数, 其是源模式和目标

可靠的中介模式间最可靠的映射。本文提出的 SPM 算法与现有的概率映射 (possibility)、概率-可靠映射 (RMedMap) 以及 falcon 映射进行了对比。

1 处理可靠映射间的不确定性

1.1 数据空间中的映射架构

数据空间中, 模式匹配和可靠映射的映射架构如图 1 所示。从连接到数据空间的数据源, 首先提取其内部图表示; 其次, 进行语法, 语义和结构匹配, 来推导出图表示之间的语义和结构关系, 将语义和结构关系合并在一起, 产生可能的中介模式; 最终, 系统计算中介模式的可靠度, 生成可靠的语义映射传递给查询。

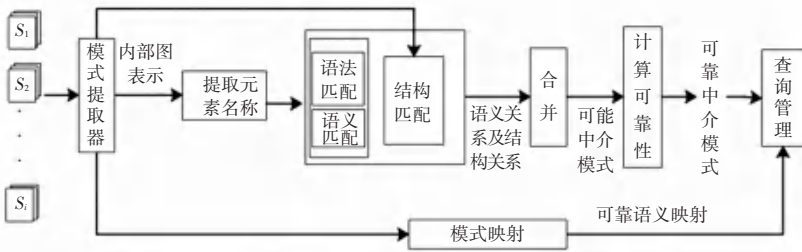


图 1 映射架构

Fig. 1 Mapping framework

1.2 可靠的模式匹配和映射原则

1.2.1 相似性计算

假设两个源模式 S_i 和 S_j , e_i 和 e_j 分别是 S_i 和 S_j 的两个非叶元素节点, 非叶元素节点是从源模式的对应图表示中收集的概念, 并非是其叶子结点。引入信息熵概念计算元素 e_i 和 e_j 的语义相似度, 式 (1):

$$sim(e_1, e_2) = \frac{d(MS(e_1, e_2)) + \delta}{IC(e_1) + IC(e_2)} \quad (1)$$

其中, $IC(c) = -\log_2 p(c)$ 是元素 e 所含的信息量, δ 是正实数。

基于元素节点 e_i 和 e_j 所处的图模式中的位置, 计算 e_i 和 e_j 之间的结构相似度^[13], 式 (2):

$$op(e_1, e_2) = \frac{d(MS(e_1, e_2))}{d(e_1) + d(e_2)} \quad (2)$$

其中, $d(e)$ 是 e 在这个模式图表示中的等级, $d(MS(e_1, e_2))$ 表示 e_1, e_2 最近的父节点的位置距离。

根据语义关系以及结构关系合并, 构建出可能的中介模式, 如对于铝电解的原料构成成分中的两个源模式 S_1 和 S_2 , 如图 2 所示。

对于源模式 S_1 和 S_2 , S_1 中的 ingredient 通常是用于氧化铝浓度分析的元素, 是关于元素的常见信息, name 可以是一个普通名称 (CommonName), 科学名称 (ScientificName), 方言 (VernacularName), raw material 是氧化铝原材料的元素, 当用户进行查询时, 计算相应的可靠度 $d(m)$ 将 S_1 中 ingredient 映射到 S_2 中的 raw materia, 得到两个模式间可能的映射为: (ingredient, raw material)、(site, reference) 和 (name, name) 等。

1.2.2 可靠度计算

考虑到一组可能的中介模式的实例 T 和给定的源模式 S_i , 其目的就是要找出 T 关于 S_i 是否可靠, 并赋予 T 一个关于 S_i 的可靠度。可靠度从两个方面来考虑, 首先以结构角度来检验 T 相对于 S_i 是否可靠, 称为结构可靠度, 若 T 的根节点在结构上是可靠的, 那么 T 相对于 S_i 则是可靠的。结构可靠度的计算方法为结构上等同于 S_i 子节点的 T 的子节点数与 S_i 的子节点数之比。 T 相对于 S 的结构可靠度公式 (3) 如下:

$$d_{T/S_i}^{struct} = \frac{|T \cap S_i|}{|S_i|} \quad (3)$$

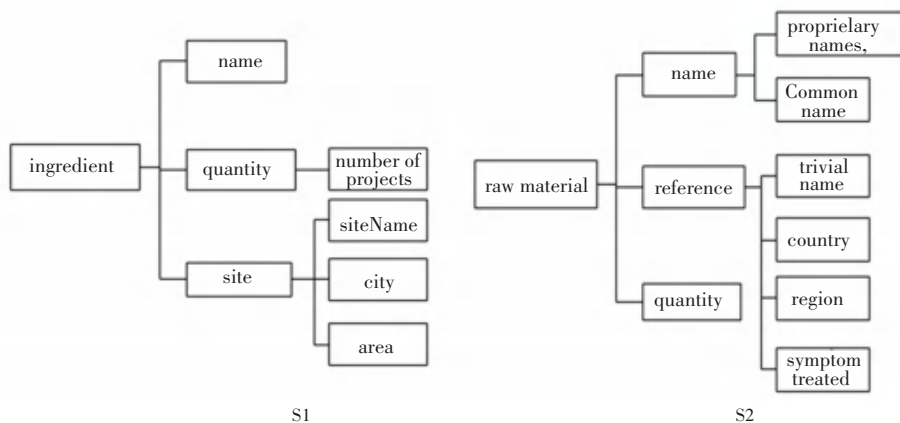


图 2 模式对比

Fig. 2 Pattern comparison

计算 T 关于 S_i 的可靠度,式(4):

$$d_{T/S_i} = \frac{\sum_{e \in T \cap S_i} d(e)p(e)}{\sum_{e \in S_i} d(e)p(e)} \quad (4)$$

其中, $d(e)$ 是元素 e 所属组元素之间的相似值, $p(e)$ 表示在中介模式中遇到元素 e 实例的概率, 是包含 e 的中介模式的数量除以中介模式的总数的

比值。

如果满足条件(1)和(2), T 对 S_i 是可靠的:

$$(1) d_{T/S_i}^{struct} > \alpha;$$

$$(2) d_{T/S_i} > \beta.$$

对于上文提到的源模式 S_1 和 S_2 , 表 1 给出了一组 5 种可能的相互不同的映射 $m_i (i = 1, \dots, 5)$ 及其相应可靠度。

表 1 可能映射及可靠度

Tab. 1 Possible mapping and reliability

编号	可能的映射	$d(m_i)$
m_1	(ingredient, raw material), (quantity, quantity), (name, name), (site, reference)	0.272
m_2	(number of projects, quantity), (name, commonname), (site, city)	0.235
m_3	(name, triaialname), (site, region)	0.210
m_4	(sitename, triivial name), (city, city)	0.176
m_5	(area, region)	0.106

2 可能性理论

语义映射的数量较大时, 阈值的选择可能成为一项不确定的任务, 并可能导致信息丢失。因此, 引入了基于可能性理论的后映射方法 (SPM), 用于管理可靠映射之间的不确定性。

可能性理论是一种专注于处理不完全信息的不确定性理论, 类似于概率论。通过使用一对称为可能性和必要性测度的对偶函数来消除概率, 而不是仅使用一个。也就是说, 事件分离的可能性程度是单个事件可能性程度的最大值。相反, 事件关联的必然性程度是单个事件的最小必然性程度。

使用最大值和最小值运算, 以及对 RmedMap 方法的弥合, 符合计算简单性的要求。在许多的真实应用中, 这种不确定性具有相当粗略和定性的性质。

但在可能性理论中, 不确定性的建模可能仍然是定性的。因此, 需要使用一个有限的完全有序的可靠性等级链, 在这里用在 $0 \sim 1$ 之间的数 λ 来表示, $\lambda_1 = 0 < \lambda_2 < \lambda_3 < \dots < \lambda_n = 1$ 。

2.1 后映射算法

构建 DSSP 的目的是为用户提供信息检索或关键字搜索等基本功能。因此, 系统产生的结果应考虑与 DSSP 连接的源中所有可用的信息。rMedMap 是一种基于可靠性的方法, 使系统能够利用源中尽可能多的可用信息, 相比之下, 概率映射是一种基于概率的方法, 考虑的可用信息较少。本文提出的后映射方法利用连接到 DSSP 的源中所有信息。SPM 映射方法是一种基于步进的方法, 其使用后一组可用信息的结果提供当前可用信息集的结果, 如图 3 所示。

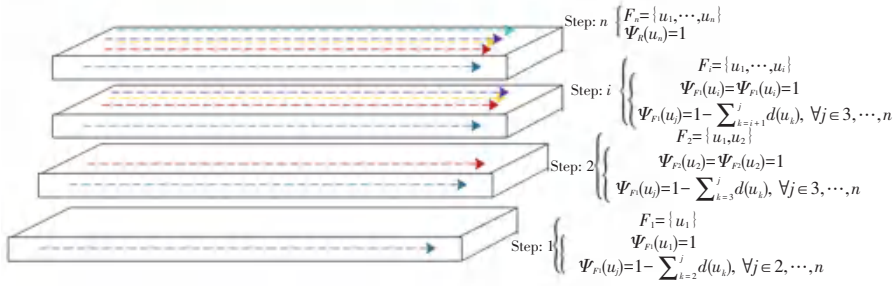


图3 步进的后映射方法

Fig. 3 SPM (stepping post mapping)

在 step1 中,系统只处理子集 F_1 及其相应的特征函数 Ψ_{F_1} , 在 step2 中进一步对子集 F_1 进行扩展, 构造出子集 F_2 及其对应的特征函数 ψ_{F_2}, ψ_{F_2} 也携带后一个特征函数 ψ_{F_1} 的信息, 以此类推。

2.2 SPM 映射过程

引入可靠映射定义, 对于源模式 S , 目标模式 T , 一个可靠的中介模式是一个 $(T, d_{T/S_i})$, T 是可能的中介模式, d_{T/S_i} 是 T 对于 S 的可靠性程度。

一个与标签 $tag(x_i, x_j)$ 相匹配的标签是一个四元组:

$$Tag(x_i, x_j, sim(x_i, x_j), op(x_i, x_j)) \quad (5)$$

$sim(x_i, x_j)$ 为 x_i 和 x_j 之间的语义相似性, $op(x_i, x_j)$ 是 x_i 和 x_j 在结构上的相似度。例如, 对于实验数据集中 ingredient 与 raw material 之间的匹配标签为:

$$tag(ingredient, raw\ material) = (ingredient, raw\ material, 0.78, =)$$

一个可靠的映射 m 是一对元组 $(m, d(m))$, m 是 S 和 T 之间的一组相互不同的一对一标签匹配结果, $d(m)$ 是映射 M 对 S 的可靠度。根据已给出的应用实例, 得到一个可靠的映射:

$$\begin{aligned} (m_1, d(m_1)) = & ((ingredient, raw\ material), 0.78, =), \\ & (quantity, quantity, 0.93, =), \\ & (name, name, 0.93, =), (site, reference, 0.5, =) \end{aligned}$$

S 和 T 之间得到的可靠的映射集 M 是一组成对的组合:

$$\{(m_i, d(m_i)), i = 1 \cdots n\}, i \neq j, m_i \neq m_j,$$

$$\sum_{i=1}^n d(m_i) = 1$$

$M = \{(m_1, d(m_1)), (m_2, d(m_2)), (m_3, d(m_3)), (m_4, d(m_4)), (m_5, d(m_5))\}$ 是一个可靠的映射集, 是源模式和目标模式之间相互独立的可靠映射集。为语义映射分配可靠度有利于克服自动生成的

多个映射之间的不确定性, 管理所得到的可靠映射集, 大多数方法采用了可靠性阈值。阈值方法会导致一个层面的不确定性。事实上, 如果语义映射是互不相同的, 那么可靠度的值取决于系统自动产生的语义映射的数量。如果语义映射的数量较高, 则可靠性会较低, 这样选择一个阈值来进行可靠映射集的选择成为了一项不确定的任务。例如, 若选择一个阈值为 0.1, 隐藏在映射 m_5 中的数据信息会丢失, 这是因为 $d(m_5) < 0.1$ 。为了解决这样的不确定性问题, 引入可能性理论, 引导系统处理所有可用的数据信息, 提出了 SPM 映射方法, 将可靠的映射集划分为可能性分布的子集。为了构造这些子集, 对得到的映射集合 M 中的元素进行降序的排序, 并且向该集合固定的增加两个可靠映射元组, $(m_{high}, d(m_{high}))$ 和 $(m_{low}, d(m_{low}))$, 使得 $d(m_{high}) = 1, d(m_{low}) = 0$ 。一个有序集合 $U = \{(u_1, d(u_1)), \dots, (u_n, d(u_n))\}$ 是一个可靠的映射集 M 通过降序得到, 并使得:

$$(1) d(u_1) > d(u_2) > \dots > d(u_n)$$

$$(2) (u_1, d(u_1)) = (m_{high}, d(m_{high}))$$

$$(3) (u_n, d(u_n)) = (m_{low}, d(m_{low}))$$

通过得到的可靠映射集, 将加入的两个可靠映射元组构成集合, 可以构建一个有序的映射集 $U = \{(u_1, d(u_1)), (u_2, d(u_2)), \dots, (u_7, d(u_7))\}$, 且满足:

$$(1) (u_1, d(u_1)) = (m_{high}, 1)$$

$$(2) (u_2, d(u_2)) = (m_1, 0.272)$$

$$(3) (u_3, d(u_3)) = (m_2, 0.235)$$

$$(4) (u_4, d(u_4)) = (m_3, 0.210)$$

$$(5) (u_5, d(u_5)) = (m_4, 0.176)$$

$$(6) (u_6, d(u_6)) = (m_5, 0.106)$$

$$(7) (u_7, d(u_7)) = (m_{low}, 0)$$

得到的映射集合 U 为一个 $n > 1$ 的有序映射集, 并且 $1 - d(u_i) = d(d(u_{n+1-i}))$, $U = \{(u_1, d(u_1)),$

$(u_2, d(u_2)), \dots, (u_7, d(u_7))\}$, 使 $d(u_1) > d(u_2) > \dots > d(u_n)$ 。定义 n 个子集 $F_i, F_i = \{u_1, u_2, \dots, u_i\}$ (n 为子集 F_i 的数目), 特征函数 ψ_{F_i} 由映射集的可靠度得到, $1 \leq i \leq n$, 对应的特征函数如下:

$$F_i = \{u_1, u_2, \dots, u_i\} \quad (6)$$

$$\uparrow F_1 = \{u_1\}$$

$$\uparrow \psi_{F_1}(u_1) = 1$$

$$\uparrow \psi_{F_1}(u_2) = 1 - d(u_2) = 0.728$$

$$\uparrow \psi_{F_1}(u_3) = 1 - d(u_2) - d(u_3) = 0.493$$

$$\uparrow \psi_{F_1}(u_4) = 1 - d(u_2) - d(u_3) - d(u_4) = 0.283$$

$$\uparrow \psi_{F_1}(u_5) = 1 - d(u_2) - d(u_3) - d(u_4) - d(u_5) = 0.107$$

$$\uparrow \psi_{F_1}(u_6) = 1 - d(u_2) - d(u_3) - d(u_4) - d(u_5) - d(u_6) = 0.001$$

$$\uparrow \psi_{F_1}(u_7) = 1 - d(u_2) - d(u_3) - d(u_4) - d(u_5) - d(u_6) - d(u_7) = 0.001$$

根据这个例子, 还可以构造可能性分布的子集 F_2, F_3, F_4, F_5, F_6 和 F_7 及其对应的特征函数。 $\psi_{F_2}, \psi_{F_3}, \psi_{F_4}, \psi_{F_5}, \psi_{F_6}$ 和 ψ_{F_7} , 迭代特征函数 ψ_{F_i} 与可靠度函数 $d(u)$ 比较, $F_i \subseteq U$ 。

设 F_i 是 U 的第 i 个子集, 使得 $u_{j_1}, u_{j_2} \in U$, 其中至少一个属于 $F_i, F_i(u_{j_1}) > F_i(u_{j_2})$, 当且仅当 $d(u_{j_1}) > d(u_{j_2})$ 。

可靠匹配算法如下:

输入: 可能的中介模式 T_1, T_2, \dots, T_m

源模式 S_i

输出: 可靠的中介模式

计算模式间语义相似度以及结构相似度, 得到标签匹配元组 tag , 合并得到可能的可靠中介模式 T

可行理论: 引入两个额外的可靠中介模式, $(m_{\text{high}}, d(m_{\text{high}})), (m_{\text{low}}, d(m_{\text{low}}))$, 使得 $d(m_{\text{high}}) = 1, d(m_{\text{low}}) = 0$ 。

开始

循环 $j = 1 \sim m$

判断 计算中介模式 T 对于源模式 S 的可靠度 $d(m)$

得到可靠映射元组 M , 根据其可靠度 $d(m_i)$ 降序排序

返回降序排序的映射集 U

迭代映射 定义 N 个子集 $F_i, F_i = \{u_1, u_2, \dots, u_i\}$, 根据式(5) 特征函数得到新可靠度, 获得降序排序后的映射子集, 映射子集不断迭代, 直到结束

结束循环

提取所有可靠映射集

结束

$$\psi_{F_i}(u_j) = \begin{cases} 1, & u_j \in F_i \\ 1 - \sum_{k=i+1}^j d(u_k), & \text{其它} \end{cases} \quad (7)$$

基于上述实例以映射子集 F_1 为例, 映射子集 F_1 相应特征函数, 式(8)。

(8)

3 实验分析

本算法使用 java 语言实现, 在 IntelliJ IDEA 平台上实现。为验证本方法的可行性, 从国内某铝厂数据平台中获取实验数据集进行实验。共选取 120 个数据集, 共 4 类, 其中有 6 个数据集词汇与结构上都相似, 34 个数据集词汇不相似, 结构上相似, 12 个数据集词汇上相似, 78 个结构上不相似数据集, 见表 2。

表 2 数据集分布

Tab. 2 Distribution of dataset

编号	类别	数据集数目
1	词汇, 结构相似	6
2	词汇上相似, 结构相似	34
3	词汇相似, 结构不相似	12
4	词汇结构都不相似	68

采用标准指标对方法进行评估, 即精度、召回率与 F -measure 方法作为映射方法的指标。 T 是由映射工具提供的非空映射, T_e 是由专家匹配器提供的非空映射。

Precision: 表示由映射工具产生的映射中正确映射的比例, 式(9)。

$$Precision = \frac{|T \cap T_e|}{|T|} \quad (9)$$

Recall: 表示由系统提取的正确映射的比例, 式(10)。

$$Recall = \frac{|T \cap T_e|}{|T_e|} \quad (10)$$

F -measure: 是建立在精度 *Precision* 和召回率

Recall 上面的一个折衷, 式(11)。

$$F - measure = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (11)$$

表 2 列出了结构与语义不同的 4 种数据集, 将本文提出的 SPM 算法与已有的算法进行比较, 结果如图 4、表 3 所示。

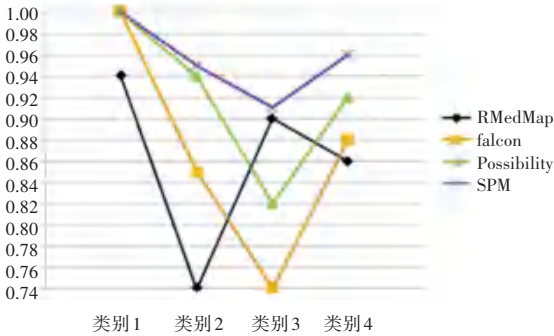


图 4 不同方法的查准率比较

Fig. 4 Comparison of precision rates of different methods

表 3 查全率与 F-measure 方法比较

Tab. 3 Comparison of recall rate and F-measure

方法	类别 1	类别 2	类别 3	类别 4	F-measure
Falcon	1.00	0.82	0.86	0.89	0.885
Possibility	1.00	0.87	0.77	0.88	0.900
RMedMap	1.00	0.76	0.78	0.85	0.855
BSP	1.00	0.85	0.79	0.88	0.914

由上述结果对比可看出, 对于词汇与结构都相似类别的数据集, 在每一种方法中的映射效果都较好, Falcon、Possibility 以及 RMedMap 方法得到的精度与召回率相对来说都比较高; 对于 F-measure 综合指标, 本文的映射方法能够比较好的得到较好精确的映射结果, 综合评价效果较好。

对于类别 2 来说, 本文的方法得到的精确率较高, 在 Recall 方面, 次于 Possibility 方法; 查全率方面, 本文的算法有待提高。对于类别 3 和类别 4 来说, 精确度在这些方法中具有优势, 召回率方面有待提高, 从 F-measure 结果来看, SPM 方法在精确度与召回率两个方面最好。

4 结束语

本文针对数据空间中动态的数据集成方式及数

据空间平台上的不确定性映射问题, 提出基于可能性理论的后映射方法, 管理独立的源模式自动提供的可靠映射。本文的方法使系统自动管理包含在自动提供的映射中的所有可用信息, 使用降序排序的方法, 得到所有的可靠映射集合, 在得到高可靠映射结果的同时得到了所有可用的信息, 但在查全率上还有待提高。

参考文献

- [1] 李玉坤, 孟小峰, 张相於. 数据空间技术研究[J]. 软件学报, 2008(8), 18-31.
- [2] 杨丹. 数据空间中基于语义的实体搜索关键技术研究[D]. 沈阳: 东北大学, 2012.
- [3] DAS SARMA A, DONG X, HALEVY A. Bootstrapping pay-as-you-go data integration systems [C]//Proceedings of the 2008 ACM SIGMOD international conference on Management of data. 2008: 861-874.
- [4] HALEVY A, FRANKLIN M, MAIER D. Principles of dataspace systems [C]//Proceedings of the twenty-fifth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems. 2006: 1-9.
- [5] SARMA A D, DONG X L, HALEVY A Y. Uncertainty in data integration and dataspace support platforms [M]. Schema Matching and Mapping. Springer, Berlin, Heidelberg, 2011: 75-108.
- [6] DO H H, RAHM E. Matching large schemas: Approaches and evaluation [J]. Information Systems, 2007, 32(6): 857-885.
- [7] MELNIK S, GARCIA - MOLINA H, RAHM E. Similarity flooding: A versatile graph matching algorithm and its application to schema matching [C]//Proceedings 18th International Conference on Data Engineering. IEEE, 2002: 117-128.
- [8] RAHM E, BERNSTEIN P A. A survey of approaches to automatic schema matching [J]. the VLDB Journal, 2001, 10(4): 334-350.
- [9] NGUYEN Q V H, NGUYEN T T, MIKLÓS Z, et al. Pay-as-you-go reconciliation in schema matching networks [C]//2014 IEEE 30th International Conference on Data Engineering. IEEE, 2014: 220-231.
- [10] 黄毅芳, 吴扬扬. 本体在个人数据空间中的应用研究 [J]. 软件时空, 2009, 8(3): 199-201.
- [11] 苏畅. 面向数据空间的实体解析方法研究 [D]. 哈尔滨: 哈尔滨工程大学, 2019.
- [12] 寇月, 申德荣, 聂铁铮, 等. S-RSM: 一种基于语义的数据空间资源搜索机制 [J]. 小型微型计算机系统, 2011(2): 265-270.
- [13] KUICHEU N C, WANG N, TCHUISSANG G N F, et al. Managing uncertain mediated schema and semantic mappings automatically in dataspace support platforms [J]. Computing and Informatics, 2013, 32(1): 175-202.