

## 基于贝叶斯 - 高斯模型的多本体映射算法

徐德智, 易晓媛, 汤 哲

(中南大学 信息科学与工程学院, 湖南 长沙 410083)

**摘 要:** 语义网中, 本体映射是解决本体异构的关键. 在面对多本体模型同时映射的情况时, 提出一种新的基于贝叶斯-高斯模型的本体映射算法 IOBGM(Improved ontology Bayesian-Gaussian model). 该方法首先将本体模型转换成贝叶斯网络模型, 接下来结合传统基于贝叶斯网络模型的本体映射方法, 采用混合高斯模型对混合节点群进行聚类判定. 最后运用了迭代映射模式, 在获得初步映射节点对相似度的情况下通过迭代模块来对其他隐性相似节点进行二次相似度判定, 从而减少结果误差. 实验结果表明, IOBGM 系统在本体查全率方面有较为突出的表现, 在多本体映射的情况下具有一定的应用优势, 其稳定性和效率能够满足实用需求.

**关键词:** 本体异构; 本体映射; 贝叶斯网络; 高斯模型; 聚类判定; 迭代映射

中图分类号: TP301.6

文献标识码: A

文章编号: 1000-7180(2017)08-0023-06

DOI:10.19304/j.cnki.issn1000-7180.2017.08.005

## An Ontology Mapping Method Based on Bayesian and Gaussian Model

XU De-zhi, YI Xiao-yuan, TANG Zhe

(Central South University, School of Information Science and Engineering, Changsha 410083, China)

**Abstract:** In Semantic Web, Ontology mapping is the key to solving ontology heterogeneous. In this paper, a new ontology model based on Bayesian-Gaussian is proposed in the face of multiple ontology mapping. First of all, the method transforms the ontology model into a Bayesian network model. Then, based on traditional Bayesian network model, the hybrid Gaussian model is used to cluster the mixed node groups. Finally, iterative mapping mode is used to reduce the error. In the case of obtaining the similarity degree of the initial mapping node, the similarity degree of other similar similarity nodes is determined by the iterative module. The experimental results show that the IOBGM system presented in this paper has outstanding performance on ontology recall rate, and has certain application advantages in the case of multi-ontology mapping. Its stability and efficiency can meet the practical requirements.

**Key words:** heterogeneous ontology; ontology mapping; bayesian network; gaussian model; clustering determination; iterative mapping

### 1 引言

语义网是现阶段网络发展的新趋势. 本体和本体映射是语义网的基础. 本体映射是从确定的知识中获取信息来推导可能的结果的过程. 本体异构导致了语义网不能整合成一个有机的整体. 为了解决本体异构的问题, 当前提出了多种本体映射的方法, 如基于概念相似度的本体映射算法, 基于现有映射的本体映射方法等.

目前, 有一些方法将贝叶斯网络模型引入到本体映射中, 如文献[1]提出的 OMEN 方法, 文献[2]提出的 BAYOWL 方法, 如文献[3]提出的 OM-OBN 方法. 但是大多数提出的方法主要针对于双本体映射的问题, 而现今语义网中多本体映射问题更为显著. 在多本体映射领域, 现今国内外研究者主要提出的方法有基于参考本体[4]的映射方法等, 但此类方法在寻找参考本体时过程复杂, 并且很多领域没有可以选择的参考本体. 因此, 本文在双本体贝叶

收稿日期: 2016-04-20; 修回日期: 2016-05-22

基金项目: 国家自然科学基金资助项目(31470028); 中南大学高校教师研究基金

斯模型映射的基础上,运用高斯模型,提出一种针对多本体映射的贝叶斯-高斯模型方法。

## 2 模型转换原理

IOBGM的第一个主要模块为模型转换.本体的定义为一个五元组  $O = \{C, H^C, R, I_O, A_O\}$  其中  $C$  表示概念名称集合,  $H^C$  表示关系集合,  $R$  表示概念间的非分类关系,  $I$  中包括了现实世界中具体的、唯一的个体即实例集合,  $A_O$  表示公理集合. 贝叶斯网络的定义如下。

### 定义 1

设贝叶斯网络模型的定义为一个三元组  $BN = \{V, A, CPT\}$ , 其中:  $V$ : 顶点集合, 代表贝叶斯网络中的随机事例的集合, 一个随机事例可以表示为一个关于  $v$  的变量  $P(V)$ 。

Ar: 弧集合, 表示贝叶斯网络中点与点之间的因果关系的集合, 即一个点确定值的概率对其他点的影响。

CPT: 条件概率表集合. 在贝叶斯网络中, 存在一个包含所有节点的概率表, 该表用以表示各个节点出现可能值得概率。

### 定义 2

在文献[5]中将本体贝叶斯网络模型的形式化定义为以下形式, 本文对其进行引用:  $OB_N = \{G = (N, E), P, I, A, \theta_{ST}\}$

$G$  中包含双元素, 即  $G = (N, E)$ . 其中,  $N$  与  $E$  分别代指网络中节点与边的集合. 对应到本体中的元素,  $N$  对应于概念集合  $C$ ,  $E$  对应于关系集合  $H^C$ 。

$P$ : 节点的属性集合. 由转换规则  $Y$  可得到关于  $P$  的变量集合,  $Y(R) = P$ 。

$I$ : 节点的实例集合. 由转换函数  $J$  可得到关于  $I$  的变量集合,  $Y(I_O) = I$ 。

$A$ : 本体贝叶斯网络模型的公理集合. 与源本体和目标本体类似,  $A$  代指本体中所有的公理集. 通过转换函数  $J$  本体模型与本体模型的公理集可实现切换, 即有  $Y(A_O) = A$ 。

$\theta_{ST}$ : 节点的相似度表集合. 将节点相似度表称为 ST(similarity table), 由此以结合 PN 与 IN 得到关于节点的相似度表集合 ST. ST 存储了源本体贝叶斯网络模型节点与其他贝叶斯网络模型中相关节点之间的相似度值。

在 IOBGM 系统中运用转换规则  $Y$ , 将本体模型中的元素分别转换成本体-贝叶斯网络模型元素, 对应关系为:  $Y(C) = N, Y(H^C) = E, Y(R) = P$ ,

$Y(I_O) = I, Y(A_O) = A$ .  $Y$  转换规则如下: 在本体 OWL 文件中:

规则一:  $\langle \text{OWL:Ontology rdf:about} = "xxx" \rangle$  标签标注本体模型的名称“O”, 贝叶斯网络模型与本体模型之间是一一对应的关系;

规则二: 本体中名称为“xxx”的概念名称集合也就是节点名称  $C$  用标签  $\langle \text{OWL:Classrdf:ID} = "xxx" \rangle$  标注;

规则三: 本体中节点间的上下位关系边集合  $H^C$  用标签  $\langle \text{rdfs:subClassOfrdf:resource} = "xxx"/ \rangle$  标注;

规则四: 非分类关系  $R$  用标签  $\langle \text{OWL:onpropertyrdf:resource} = "xxx"/ \rangle$  标注得到贝叶斯网络中的  $P$ ;

规则五: 实例集合  $I_O$  用标签  $\langle \text{rdf:Descriptionrdf:ID} = "xxx" \rangle$  标注;

规则六: 本体中所有完整的 OWL 文件片段都属于公理类. 因此公理集  $A_O$  可以通过规则二、三、四来实现。

## 3 多本体同时映射相似度表的建立

本节主要讨论的是在多个本体模型同时进行映射时, 如何建立合理的相似度表. IOBGM 针对多本体节点相似度表构建稍作调整, 引入混合高斯模型. 以下分析中, 仅就节点的文本信息的描述特征进行详细介绍。

在多本体贝叶斯网络模型同时映射时, 在得到每个模型中相关节点关于标签及其注释的向量矩阵  $\vec{V}_{jk} = (s_{1k}, s_{2k}, \dots, s_{jk})$  后, 将其转化为  $k$  个  $j$  维的列向量  $\vec{X} = (x_1, x_2, \dots, x_k)$ , 使用 K-means 聚类算法求得来源占比  $a_j$ , 再代入到混合高斯模型中获得其概率密度, 由混合高斯模型运算器选取高密度节点对作为映射节点, 从而确定映射节点对之间的相似度表. 文献[6]中提及, 贝叶斯网络模型通过聚类来进行节点分类, IOBGM 的设计与该方法类似. 节点相似度表建立这一阶段的流程, 如图 1 所示。

### 3.1 节点基于矩阵的描述形式

IOBGM 中仅就节点的文本信息的描述特征对其进行分析<sup>[7-8]</sup>。

本体中各节点的所包含的实体的文本信息描述信息通常用标签及注释的方式进行说明, 当描述信息类似时, 那么实体之间有可能存在某些联系, 也可以说包含该实体的节点具备某种关系. 将本体的每

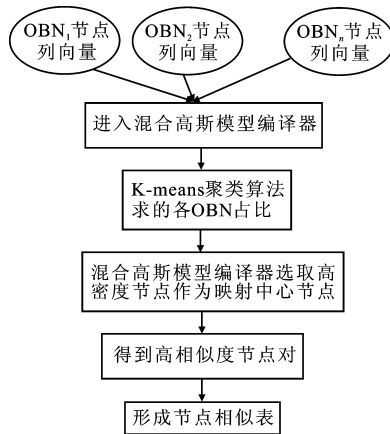


图 1 混合高斯模型编译器内部运作图

个节点所包含的实体描述信息表示成向量的形式有： $\vec{V}_{jk} = (s_{1k}, s_{2k}, \dots, s_{jk})$ 。

其中  $j$  为该节点下的不重复标签的个数,  $k$  为该关键词相关的注释出现的个数,  $j, k \in (0, +\infty)$ ,  $s_{jk}$  为该节点描述信息同时包含  $j$  标签及  $k$  关键词的实体出现的次数,  $j, k$  的设置由专家判定及机器经验给出。当该节点中不包括  $j, k$  对应的实体时  $s_{jk}$  为 0, 反之则为 1。从而形成了一个  $j \times k$  的矩阵, 如表 1 所示。进而将计算节点相似度的问题转移到计算矩阵相似度上来。

表 1 节点基于描述信息的矩阵形式

	$k_1$	$k_2$	$k_3$	...	$K_m$
$J_1$	1	0	1	...	0
$J_2$	0	0	0	...	0
$J_3$	1	1	0	...	1
...	...	...	...	...	...
$J_n$	0	1	1	...	0

表 1 中, 有  $n \times m$  矩阵, 表示了节点  $\vec{V}_{jk} = (s_{1k}, s_{2k}, \dots, s_{jk})$  的矩阵形式, 其中行向量  $J$  代指不重复标签的个数, 列向量  $k$  代指与关键词相关的注释出现的个数, “0” 代表该节点不同时具备该标签和关键词注释, “1” 反之。

### 3.2 高斯模型

在得到各个本体节点列向量之后, 使用 K-means 算法求出各个本体的来源占比。

在多个本体模型同时进行映射时, 用混合高斯模型来计算节点相似度。给出高斯模型及混合高斯模型的参数计算分析。

高斯模型(SGM)是有向无环模型的一种, 是贝叶斯网络模型的拓展, 可避免节点概率重叠。将高斯模型定义为以式子, 多维向量  $X$  服从高斯分布时,

它的概率密度函数为:

$$N(x; u, c) = 1 / \sqrt{(2\pi)^d |C|} \exp \left[ -\frac{1}{2} (x - u)^T C^{-1} (x - u) \right] \quad (1)$$

式中,  $d$  是向量维度,  $\mu$  是列向量的数学期望,  $C$  为列向量的协方差矩阵,  $|C|$  是  $C$  的行列式,  $C^{-1}$  表示  $C$  的逆矩阵,  $(x - u)^T$  是  $(x - u)$  的转置矩阵。

混合高斯模型(GMM)<sup>[9]</sup> 是单一高斯概率密度函数的延伸。GMM 是一种聚类算法模型, 即在只有样本点, 不知道样本分类的情况下, 计算出模型参数。例如: 有  $n$  个观察向量  $\vec{X} = (x_1, x_2, \dots, x_k)$ , 构成  $n$  个  $j$  维的列向量, 由于数据的复杂程度加大, 单一的高斯密度函数无法描述多类别数据点。因此引入新的模式, 即假设每个矩阵均来源一个单一高斯分布, 先忽略其分布的相关参数问题, 从而有一组关于数据点的高斯模型, 其中任意  $x_i$  来哪个高斯模型暂且不考虑, 单个模型在混合中占的占比  $\alpha_j$  未知, 将所有的  $x_i$  合并后得到的分布就是混合高斯分布。

这些数据的概率分布密度函数可以通过加权函数表示, 设 GMM 共有  $n$  个 SGM, 可通过采样样本集来估计 GMM:  $\theta = (SGM_1, SGM_2, \dots, SGM_n)$ , 样本  $X$  的概率公式为:

$$P(x) = \sum_{i=1}^n \alpha_i N_j(x_i; u_i, C_i), \sum_{i=1}^n \alpha_i = 1 \quad (2)$$

$\alpha_i$  为混合高斯模型中单个高斯模型的占比参数, 在  $n$  个  $\vec{X} = (x_1, x_2, \dots, x_k)$  中各个数据点分类已知的情况下,  $\alpha_i = N_i/n$ 。在建立起模型并求出各个参数以后, 根据需要给出模型阈值  $T$ , 由此来确定选择出来的相似度范围。

下面给出混合高斯模型编译器实现算法伪代码。算法 G-V:

```

Input: 向量空间组  $\vec{X} = (x_1, x_2, \dots, x_k)$ 
Output: 高相似度节点对  $A(p) = (P_1, P_2, P_3, \dots, P_n)$ 
For Each  $x \in (x_1, x_2, \dots, x_k)$ 
     $N_j = N(x, u, c)$ 
     $\alpha_j = N_j/n$ 
     $p_j(x) = P(x)$ 
    For Each  $x \in p_j(x)$ 
         $p_j(x) > T$ 
         $p_j(x) \in A(p)$ 
    EndForEach
EndForEach

```

该算法由两重循环构成, 外循环的步骤是输入

向量空间组,根据高斯和混合高斯模型公式,求出其概率密度和概率。内循环是对于每一个节点进行判定,当其在混合高斯模型中的概率大于事先指定的阈值时,将其加入到高相似度节点对的范围之内。经过该算法后,将选出来的多个节点加入到高相似度节点对数组中,进行下一步映射表的建立。

### 3.3 映射相似度表的建立

通过以上两个步骤,得到了高密度相似度节点对。在进行映射时,本文根据图 2 多个本体节点相似度示例关系图,将高密度节点对的相似度通过二维数组来表示,从而形成多个关于节点坐标的相似度矩阵。

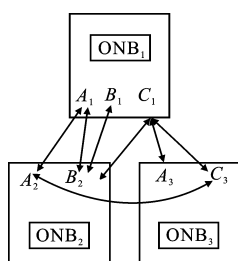


图 2 多个本体节点相似度示例关系图

为了更好的表述,假设多个本体分别为  $ONB_1$ 、 $ONB_2$ 、 $ONB_3$ ,取第一个本体和第二个本体做详细分析。其中  $ONB_1$  的节点集合为  $N_1 = \{A_1, B_1, C_1\}$ ,  $ONB_2$  的节点集合为  $N_2 = \{A_2, B_2\}$ ,其中  $ONB_1$  任一点都可能与  $ONB_2$  中的任一点存在映射关系,取相似度值最高的节点对,从而有相似度矩阵  $SM$  (The similarity matrix) 表 2。

表 2 相似度矩阵  $SM$

	$A_1$	$B_1$	$C_1$
$A_2$	$\text{sim}_{(A_1, A_2)}$	$\text{sim}_{(B_1, A_2)}$	$\text{sim}_{(C_1, A_2)}$
$B_2$	$\text{sim}_{(A_1, B_2)}$	$\text{sim}_{(B_1, B_2)}$	$\text{sim}_{(C_1, B_2)}$

从表 2 中可以看到本体中所有的节点的相似度。如果需要  $ONB_1$  中的节点相似度,则只需将表 2 做纵向分析,从而得到关于  $ONB_1$  的节点相似度表

$$\theta_{STA1} = \{\text{Sim}(A_1, A_2), \text{Sim}(A_1, B_2)\},$$

$$\theta_{STB1} = \{\text{Sim}(B_1, A_2), \text{Sim}(B_1, B_2)\},$$

$$\theta_{STC1} = \{\text{Sim}(C_1, A_2), \text{Sim}(C_1, B_2)\}.$$

如上示例,将多个本体的节点进行分析,为任意两组有关系的本体节点建立相似度矩阵。由此,本文将多个本体映射相似度表进行了综合筛选,得到了多本体的多个相似度表集合。

## 4 相似度迭代映射推理

在两个模块的共同作用下得到贝叶斯-高斯网

络模块,称此时的模块为初始状态。初始状态经过迭代映射推理模块时,基于相似度节点表中的节点对进行二次推理映射,继续更新相似度表。即该模块首先从相似度表中寻找相似度突出的且未建立映射关系的节点对,根据所含的概念建立映射关系;利用该节点对所建立的映射关系进行第二阶段推理映射,以此反复直至没有新的映射关系被发现为止。

在 IOBGM 中采取基于相似度传播的迭代映射方式。如果两个节点一一对应具备映射关系,则它们周围的节点,包括其 Parent 类、Children 类、Predicate 类均具备相互对应的映射可能性。

相似度传播算法的过程如下,在文献[10]中对通过吸收临近信息来进行逻辑判定有详细描述。判断相似度传播算法大致分为三个阶段。

图 3 是两个已存在映射关系的节点对  $(a, b)$ 。首先是编译器基于前文所述的确定映射关系的初始相似节点对以及其临近关系构建一个节点连接图,如图 4 所示。

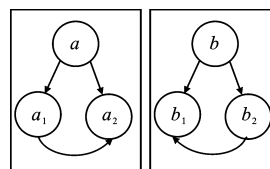


图 3 相似度传播算法节点对

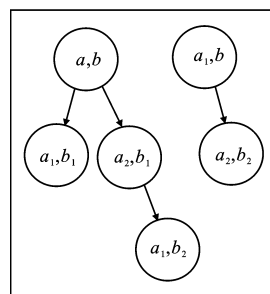


图 4 相似度传播算法节点连接图

接下来为每条边构建一个逆传播线路,定义一个以节点之间边为依据的传播系数  $\varphi$ ,其规则是认为相似度以边的多少为标准,均匀的分配给临近点,如图 5。

$$\varphi[(a_1, b_1), (a, b)] = \varphi[(a_2, b_1), (a, b)] = 1/2$$

$$\varphi[(a_1, b_2), (a_2, b_1)] = 1;$$

最后,对于该节点连接图所示的节点都有可能存在映射关系,它们的相似度在原有前文提及的相似度基础上再加上由初始相似节点对传递而来的那部分相似度,在完成所有传递后所得的值除以最大

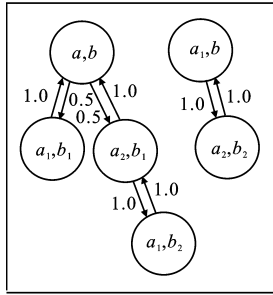


图 5 相似度传播算法逆传播线路

传递相似度值,得到最终结果,如图 6 所示.

$a, b$	1.0
$a_1, b_1$	0.91
$a_1, b_2$	0.69
$a_2, b_1$	0.39
$a_1, b$	0.33
$a_2, b_2$	0.33

图 6 相似度传播算法总相似度

在 IOBGM 系统中的相似度传播步骤如下:

1、设在前面两个模块下获得了  $OB_{N_1}$  与  $OB_{N_2}$  中的一对映射节点对  $(a, b)$ ; 将  $a$  和  $b$  临近的点设为  $(a', b')$ ;

2、然后判断依据相似度传播算法计算  $(a', b')$  获得相似度后将其存储于相似度表中用以更新;

3、以同样的方法对  $(a', b')$  的临近节点对进行迭代推理,并更新相似度表.

以上提出的方法只修改了映射节点对其周围节点相似度值,而与该节点不相邻的点则不在考虑范围.因此,减少了推理计算量,从而减少了运算时间.且由于迭代模块的存在,也避免了因信息丢失而导致结果的不准确.

通过以上三个模块来对本体-贝叶斯网络模型的节点进行相似度匹配,从而完成节点相似度表的建立,根据相似度值来判断本体之间是否具备相似的可能.

## 5 实验与分析

### 5.1 实验环境

在此选取 OAEI 竞赛中专用测试集 Benchmark (2011)来对系统进行测试.该测试集包含 110 个测试用例,大致分为以下的五类,其中 101 号本体是参

考本体,用于其他各本体与参考本体进行类比映射,101~104 号用例要么完全一样要么完全相异.201~210 号用例修改了语言学特征但未修改结构特征,221~247 号用例语言学特征未变而结构发生了变化.248~266 号用例语言学特征以及结构特征均进行了修改.301~304 号用例组采用了现实世界的用例,以测试系统实际使用价值.

由于本次实验主要针对多本体同时映射,而系统的测试集主要是双本体,因此在实验过程中,将输入本体改为输入多个测试本体.对 # 101~104、# 201~202(共十个小类)、# 221~233(共十个小类)、# 248~259(共五十个小类)、# 301~304 号本体分五组,每组内部本体元素打散后进行同时映射,以验证在多本体环境下系统的效率.

### 5.2 系统结构图

根据上述提出的思路,IOBGM 系统结构图如图 7 所示.

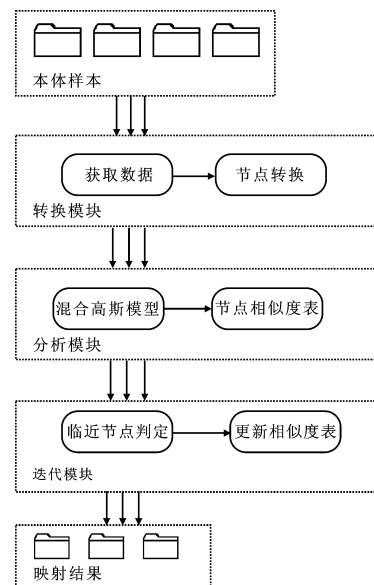


图 7 系统结构图

### 5.3 实验结果及分析

本实验所用的评估方法是现今对本体映射系统测试通用的评判标准,即通过查全率、查准率以及 F-measure<sup>[11]</sup> 对系统进行评估.其计算公式如下:

$$\text{查全率: } \text{Recall}(R, I) = \frac{|R \cap I|}{|R|} \quad (3)$$

$$\text{查准率: } \text{precision}(R, I) = \frac{|R \cap I|}{|I|} \quad (4)$$

$$F\text{-measure}(R, I) = \frac{2 \times \text{pre}(R, I) \times \text{rec}(R, I)}{\text{pre}(R, I) + \text{rec}(R, I)} \quad (5)$$

式中,  $R$  为参考结果,  $I$  为系统 IOBGM 的测试结果.

实验分析过程如下,表 3 为一般双本体映射实验数据,表 4 为多本体映射实验数据,每组得到的结果为映射后的平均值. 本次实验选择以双本体映射为基础,提出多本体的映射思路,并进行实验验证. 从实验结果来看,IOBGM 系统在多本体同时映射时其查全率以及查准率已经可以达到基准要求. 在查全率上,多本体映射比双本体映射有明显上升. 说明该方法对于多本体映射有较好的结果. 在映射时间方面,该系统对应的多本体同时映射,其用时明显少于双本体分别映射<sup>[12-13]</sup>.

表 3 IOBGM 在双本体中的实验结果

Test	Exist	Found	Found	Correct	Recall	Pre.	F-m.
101	97	97	97	97	1.000	1.000	1.000
103	97	97	97	97	1.000	1.000	1.000
104	97	59	59	59	0.608	1.000	0.756
201-2	97	70	70	70	0.722	1.000	0.838
201-4	97	94	92	92	0.969	0.979	0.974
221	97	96	80	80	0.928	1.000	0.963
248	97	56	56	56	0.577	1.000	0.732
301	58	58	58	58	1.000	1.000	1.000

表 4 IOBGM 在多本体映射分组实验结果

Test	Recall	Pre.	F-m.
101-104	1.000	1.000	1.000
201-202	0.915	0.939	0.927
221-233	0.931	0.904	0.917
248-259	0.904	0.819	0.859
301-304	1.000	1.000	1.000
AVE	0.918	0.863	0.890

## 6 结束语

本文提出一种基于贝叶斯高斯的多本体映射方法,将高斯模型引入本体映射中去. 现阶段研究本体映射主要集中在两个本体之间,多本体映射的方法还比较少,且没有统一的测试集,故本文提出的系统无法有效的同其他系统进行比较.

目前对于高斯模型的利用普遍聚集在视频图像处理方面,本文探索式的将其用于本体映射领域. 在查准率方面虽然有一定的不足,但因高斯模型其独特的聚类能力,使得系统在查全率方面提升显著,这也是新方向的一次探索. 未来在多本体映射方面,还可以进一步优化混合高斯模型,以保证查全率的同时提高查准率.

## 参考文献:

- [1] Mitra P, Noy N F, Jaiswal A R. Omen: a probabilistic ontology mapping tool[C]//International Semantic Web Conference. Berlin, Heidelberg, Springer, 2005: 537-547.
- [2] Pan R, Ding Z, Yu Y, et al. A bayesian network approach to ontology mapping[C]//International Semantic Web Conference. Berlin, Heidelberg, Springer, 2005: 563-577.
- [3] 张凌宇, 陈淑鑫. 基于本体贝叶斯网络模型的本体映射方法研究[J]. 科技通报, 2016, 31(1): 133-137.
- [4] Dragut E, Lawrence R. Composing mappings between schemas using a reference ontology[C]//OTM Confederated International Conferences. Berlin, Heidelberg, Springer, 2004: 783-800.
- [5] Caterina Caracciolo, Jtrme Euzenat, Laura Hollink, et al. Results of the ontology alignment evaluation initiative 2008 [C]//The Third International Workshop on Ontology Matching(OM2008). Karlsruhe, Germany, 2008.
- [6] Babacan S D, Molina R, Katsaggelos A K. Variational bayesian super resolution[J]. IEEE Transactions on Image Processing, 2011, 20(4): 984-999.
- [7] Kalfoglou Y, Schorlemmer M. IF-Map: an ontology-mapping method based on information-flow theory [M]. Berlin, Heidelberg, Springer, 2003: 98-127.
- [8] 邹黎君, 凌兴宏, 姚望舒. 一种基于特征信息组合的本体匹配流程[J]. 计算机工程, 2013, 39(9): 271-276.
- [9] Lee D S, Hull J J, Erol B. A bayesian framework for gaussian mixture background modeling [C]//Image Processing, 2003, ICIP 2003. Proceedings. 2003 International Conference on. Barcelona, Spain, IEEE, 2003(3): III-973-6. 2.
- [10] Borgida A, Serafini L. Distributed description logics: Assimilating information from peer sources[M]. Berlin, Heidelberg, Springer, 2003: 153-184.
- [11] Ehrig M, Euzenat J. Relaxed precision and recall for ontology matching[C]//Proc. K-Cap 2005 workshop on Integrating ontology. Banff, Canada, 2005: 25-32.
- [12] Christmas J, Everson R. Robust autoregression: student-t innovations using variational Bayes[J]. IEEE Transactions on Signal Processing, 2011, 59(1): 48-57.
- [13] Hamamoto K, Kitamura A, Taguchi S, et al. Defect cause search support system using ontology and bayesian network in liquid crystal display manufacturing process[J]. Procedia Computer Science, 2016(96): 859-868.

(下转第 32 页)

从表 3 可以看出,使用加权与差分融合参数大大提高声纹识别率.这说明由 $[\Delta\text{MFCC};\text{WMFCC}]$ 得到的融合矩阵可以分析声纹信号中帧与帧之间的短时动态特性,还能有效增强个人声纹特性上的分辨率,因此本文获取的融合参数具有更高的识别率和稳定性.

## 5 结束语

本文的实验数据和表格都是在实验室环境下基于 matlab 的平台实现的,利用多个说话人建立声纹样本,利用 VQ 的声纹识别模型,在不同类型的码本和码本容量下,采用单一特征参数和融合特征参数进行多个码本容量的实验参数.在特征参数的计算上,本文采用增减阶数法计算特征参数中每个倒谱特征参数对声纹识别系统正确率的重要度,选择倒谱特征矩阵中对声纹识别重要度高的组成新的融合特征矩阵.最后,根据实验可得出,本文获取的融合特征参数可以使得识别率提升,在以后声纹识别中将具有更好的优化与应用.

## 参考文献:

- [1] Garg A, Sharma P. Survey on acoustic modeling and feature extraction for speech recognition[C]//Computing for Sustainable Global Development (INDIA-Com), 2016 3rd International Conference on. New Delhi, India, IEEE, 2016; 2291-2295.
- [2] 孙冬梅, 裴正定. 生物特征识别技术综述[J]. 电子学报, 2001, 29(12): 1744-1748.
- [3] Shabani S, Norouzi Y. Speech recognition using prin-

cipal components analysis and neural networks[C]//Intelligent Systems (IS), 2016 IEEE 8th International Conference on. Subotica, Serbia, IEEE, 2016; 90-95.

- [4] 胡政权, 曾毓敏, 宗原, 等. 说话人识别中 MFCC 参数提取的改进[J]. 计算机工程与应用, 2014, 50(7): 217-220.
- [5] 柯晶晶, 周萍, 景新幸, 等. 差分和加权 Mel 倒谱混合参数应用于说话人识别[J]. 微电子学与计算机, 2014, 31(9): 88-91.
- [6] 申小虎, 万荣春, 张新野. 一种改进动态特征参数的话者语音识别系统[J]. 计算机仿真, 2015, 32(4): 154-158.
- [7] MarkoviC D, Popat J, Antonacci F, et al. An informed separation algorithm based on sound field mapping for speech recognition systems[C]//Acoustic Signal Enhancement (IWAENC), 2016 IEEE International Workshop on. Xi'an, China, IEEE, 2016; 1-5.
- [8] 胡航. 现代语音信号处理[M]. 北京: 电子工业出版社, 2014.
- [9] 张雪英. 数字语音处理及 MATLAB 仿真[M]. 北京: 电子工业出版社, 2012.

## 作者简介:

郑凯鹏 男, (1992-), 硕士研究生. 研究方向为语音识别. E-mail: 1049771392@qq.com.

周萍 女, (1961-), 硕士, 教授. 研究方向为语音识别与智能控制.

张上鑫 男, (1992-), 硕士研究生. 研究方向为语音识别.

柯晶晶 女, (1988-), 硕士研究生. 研究方向为语音信号处理和说话人识别技术语音识别.

(上接第 28 页)

## 作者简介:

徐德智 男, (1963-), 博士后, 教授. 研究方向为 Web 计算、语义数据库.

易晓媛(通讯作者) 女, (1991-), 硕士研究生. 研究方向为语义网、本体映射. E-mail: 624906884@qq.com.

汤哲 男, (1977-), 副教授. 研究方向为人工智能.