

## 基于证据理论的上下文本体建模以及不确定性推理方法

李艳娜 乔秀全 李晓峰

(北京邮电大学网络与交换技术国家重点实验室 北京 100876)

**摘要:** 不确定上下文信息的表示与推理是研究和开发上下文感知系统的重点和难点。该文首次将证据理论和本体相结合,提出了基于证据理论的不确定性上下文本体建模方法,并对证据组合规则进行了修改,不仅解决了证据理论在高度证据冲突时的局限性,而且使得该推理模型具有自适应性,设计并实现了不确定上下文推理算法。最后,通过在原型系统中的医疗监护和诊断应用,验证了该方法的可行性、合理性和有效性。

**关键词:** 上下文建模; 不确定性推理; 证据理论; 本体

**中图分类号:** TP391

**文献标识码:** A

**文章编号:** 1009-5896(2010)08-1806-06

**DOI:** 10.3724/SP.J.1146.2009.01015

## An Uncertain Context Ontology Modeling and Reasoning Approach Based on D-S Theory

Li Yan-na Qiao Xiu-quan Li Xiao-feng

(State Key Laboratory of Networking and Switching Technology,

Beijing University of Posts and Telecommunications, Beijing 100876, China)

**Abstract:** The modeling and reasoning with uncertain context information is one of key problems of building context-aware applications in pervasive computing environment. In this paper, a method of modeling uncertain contexts is proposed by combining the D-S theory with ontology model. And also an improved evidences combination rule is presented not only to solve the D-S theory limitations in the case of high-degree conflict evidences but also to make the reasoning have self-adaptability. Then the uncertain reasoning algorithm is designed and implemented. Finally, the feasibility, rationality and effectiveness of this approach are verified through an e-health use case of the prototype system.

**Key words:** Context modeling; Uncertain reasoning; Dempster-Shafer (D-S) theory; Ontology

### 1 引言

随着传感器、无线通信和计算技术的快速发展和普及,普适计算时代正在来临。普适计算要求计算设备能够感知用户所处的上下文及其变化,并做出相应的反应,其中,上下文信息的采集和处理具有非常重要的作用。然而,一方面,由于上下文信息模型描述存在异构性,机器对同一上下文可能有不同理解,导致很难实现知识的共享和重用,使得基于知识的推理更加困难;另一方面,由于硬件设备的局限和网络传输的问题,采集的上下文信息可能是有噪声的或不完备的,另外,从不确定的感知数据推理出的高层上下文信息也存在一定的不确定性。因此,如何形式化的描述这些上下文信息,

并且能用不确定性的知识理论进行推理、判断和决策变得尤为重要。

目前,基于贝叶斯理论的不确定性上下文建模和推理方法受到研究人员的重视, Ye, Qiao, Ko Kwang-Eun等人<sup>[1-3]</sup>提出通过添加概率标签,扩展本体,用概率表示不确定上下文,并通过构建贝叶斯网络对上下文信息进行推理。但是贝叶斯理论本身存在缺点<sup>[4]</sup>:需要收集所有要求的先验条件概率和联合概率;要求一切可能的结果都不相交;概率公式的精度依赖于完全假设集的有效性等。而证据理论是扩展了的贝叶斯理论,并克服了它的缺点,处理不确定性的能力更强。证据理论处理不确定性类似于人类的推理流程,使用信任函数和似然函数,把不知道和不确定区分开。Thomas<sup>[5]</sup>利用语义网中确定的知识、数据、推理机制、业务等能力,将D-S理论数据融合应用到专家系统中。文献[6]针对经典的证据理论没有考虑动态多源证据觉察上下文信息的可靠性、时效性和独立性因素,对证据理论进行

2009-07-17 收到, 2009-12-01 改回

国家自然科学基金(60802034, 60672122), 高等学校博士学科点专项科研基金(20070013026)和北京市科技新星计划(2008B50)资助课题  
通信作者: 李艳娜 liyanna815@gmail.com

了扩展与改进。Lu等人<sup>[7]</sup>使用D-S证据理论实现分组决策,并提出了加权的证据组合规则。LIU P<sup>[8]</sup>将证据理论应用到普适计算环境中,提出了基于证据理论组合规则的推理机制,并设计了推理机。但是以上方法只是单纯地研究基于证据理论的数据融合,基本上是直接基于采集到的数据进行融合,无法利用一些隐含的衍生信息进行推理;另外,现有模型的自适应性比较差,难以根据上下文信息的变化做出相应的调整。而基于本体的上下文感知研究,可以支持基于本体和用户自定义规则的确定性推理,但难以处理不确定性推理。因此,目前缺乏一种将本体建模和证据理论相结合的机制,既能解决上下文信息在异构系统间的知识/概念共享,实现基于本体/规则的确定性推理,又能兼顾上下文信息中存在的不确定性情况,支持不确定性推理。

本文首次将证据理论和本体建模相结合,提出了基于证据理论的上下文本体建模及不确定性推理方法(Uncertain Context Ontology Modeling and Reasoning Approach based on D-S theory, UCOMRADS),该方法既能够反映上下文信息的结构和它们之间的关系,又可以进行本体推理、规则推理和不确定性推理。本文通过建立不确定本体概念模型,综合证据距离和修正信度系数,改进了证据组合公式,使得推理模型具有自适应性,设计并实现了上下文信息的不确定推理算法,并实现了相应的原型系统(Context-Aware System supporting D-S Reasoning, CASDSR)。实验表明,本文提出的方法可以方便地对不确定性上下文信息进行本体建模,推理具有自适应性,有效地提高了不确定性推理的准确率。

本文结构安排如下:第2节提出了基于证据理论的不确定上下文本体概念模型,第3节论述了基于证据理论本体模型的不确定性推理算法,并在原型系统中进行了实验仿真;第4节对本文的工作进行了总结。

## 2 上下文建模方法

上下文建模的目的是刻画用户所处的环境或情形,在构建上下文感知系统中具有极其重要的作用,目前,上下文建模方法主要有:键值对模型、标记配置模型、图模型、面向对象模型、基于逻辑模型、本体模型等。

### 2.1 基本的本体模型

上下文信息反映了事物的属性或者状态,为了支持上下文信息在异构网络中的共享和重用,避免上下文信息的歧义性,以及支持上下文信息的推理

和机器对上下文信息的理解,本文选择语义网中的本体对上下文进行建模。本体指的是共享概念模型的明确的形式化规范说明,用于描述概念以及概念之间的关系。本体模型具有知识共享、支持逻辑推理、便于知识重用等优点,应用十分广泛。目前应用较多的本体描述语言是W3C的Web Ontology Language(OWL),它把上下文表示成一阶谓词形式,即Predicate(Subject, Object)。例如John在浴室可以表示为Location(John, bathroom)。

### 2.2 基于证据理论的本体模型

D-S理论<sup>[9]</sup>是Dempster在1968年提出的,他的学生Shafer在1971年进行了改进。D-S理论把人们所感兴趣的信息划分到一个个识别框架 $\Theta$ 中,对于任何一个属于 $\Theta$ 的子集 $A$ ,如果满足 $m(\Phi) = 0$ ,  $\sum_{A \subseteq \Theta} m(A) = 1$ ,则称 $m$ 为 $\Theta$ 上基本信度分配(Basic Probability Assignment, BPA)函数。为了表示对命题的信任程度,D-S理论引入信任函数Bel的概念,它与BPA的关系满足 $Bel(A) = \sum_{B \subseteq A} m(B)$ 。

对于信任的合成,设 $Bel_1$ 和 $Bel_2$ 是同一识别框架 $\Theta$ 上的两个信度函数, $m_1$ ,  $m_2$ 分别是对应的基本可信度分配, $A_i$ 和 $B_j$ 为焦点,则两个信度函数的合成法则为

$$m(A) = m_1 \oplus m_2(A) = \frac{\sum_{A_i \cap B_j = A} m_1(A_i) m_2(B_j)}{1 - \sum_{A_i \cap B_j \neq \phi} m_1(A_i) m_2(B_j)} \quad (1)$$

$m(A)$ 反映了 $m_1$ 和 $m_2$ 对应的两个证据对命题A的联合支持程度。 $k = \sum_{A_i \cap B_j \neq \phi} m_1(A_i) m_2(B_j)$ ,反映

了证据之间的冲突程度。当 $k = 0$ 时,称为完全不冲突;当 $0 < k < 1$ 时,称为非完全冲突; $k = 1$ 时,称为全冲突。

证据理论不能很好地反映上下文信息的结构和它们之间的关系,很大程度上限制了其应用,必须借助一种有效的上下文信息表示方法,才能发挥出它的优越性。证据理论使用证据识别框架 $E$ 和结论识别框架 $\Theta$ 来划分所感兴趣的信息,而上下文的本体建模把上下文分为两层:低层采集的上下文和高层推理出的上下文,自然而然地可以把两者关联起来,建立基于证据理论的本体模型。

基于证据理论的不确定性上下文信息的概念模型如图1所示,在基本的本体模型上添加两个类:表示结论的类Conclusions和表示证据的类Evidences,分别表示 $\Theta$ 和 $E$ ,映射到本体层次结构中就是高层推理上下文信息和低层感知上下文信息。为了表示

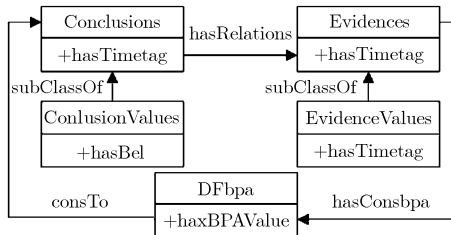


图1 证据理论的概念模型

上下文信息的新鲜度，添加数据属性hasTimetag，标识获取该上下文信息的时间。Conclusions类有子类ConclusionValue，这些子类列举出了Conclusions的所有可能取值或取值集合。同样地，Evidences有子类EvidenceValue，列举出Evidences的所有可能取值或取值集合。普适计算环境中，某个结论上下文信息可能是另一个结论的证据上下文信息，因此为了适应普适环境中上下文的动态性、多样性、不确定性，为ConclusionValue和EvidenceValue新添加数据属性hasBel，用于描述该取值的概率。在证据理论中，用信任度和BPA来表示某个事件的支持程度，因此需要在确定性的上下文模型中添加表示信任依赖关系及其概率，添加新类DFbpa，用于描述结论对证据的依赖关系，它有一个数据属性hasBPAValue，其取值即为证据理论的基本概率分配值m。为了使该不确定模型不依赖于某一具体的上下文感知应用，类DFbpa还要有两个对象属性consTo和hasConsbpa，分别描述该基本概率分配值所属的证据和结论。

上述的本体概念模型具有普遍性，只要在基本的本体模型基础上，添加该概念模型即可表示不确定性上下文信息。

### 3 上下文推理

上下文推理是普适计算的一项核心技术，是解决不确定性推理问题的关键。上下文推理的实现和上下文模型关系很大，根据本文提出的本体模型方法，上下文推理部分可以采用3种类型的推理机制，即：本体推理，规则推理和证据理论推理。

#### 3.1 基于规则/本体的推理

基于规则的推理方法是将已有的事实同预定义的规则进行匹配而产生出新的知识。例如，当传感器采集到John当前位于卧室内，室内的窗帘关着，灯光强度黑暗，则可以推断出John正在睡觉。本体的推理过程与基于规则的推理过程类似，只是使用的规则是由OWL语言本身定义的，如对称属性SymmetricProperty，传递属性TransitiveProperty等等，以便获得隐含在显式定义和声明中的知识。

#### 3.2 基于证据理论的推理

证据理论推理中，证据集和结论集分别用多个辨别框表示，合成方法提供了证据的层次推理模型，不确定性是依照D-S理论的组合规则逐层传递的，这种层次模型的建立，使得追加证据可以在原来证据的融合基础上进行，大大提高了系统的证据融合效率。

**定义** 信任结构 给定一个识别框架 $\Theta$ ，证据空间 $E$ ，映射 $F: E \rightarrow \Theta$ 以及基本信度分配值 $m$ ，其中 $\Theta$ 和 $E$ 均为有限集，则4元组 $D = \langle \Theta, E, F, m \rangle$ 称为一个信任结构。

UCOMRADS就是在基本的本体模型基础上，依据本文给出的不确定本体概念模型，构建信任结构 $D = \langle \Theta, E, F, m \rangle$ ，进而创建不确定性上下文信息的本体实例，最后利用证据组合规则逐层推理的过程。其中本体模型和信任结构的映射关系如图2所示。

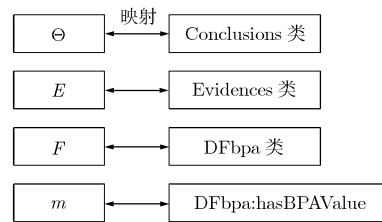


图2 信任结构和本体概念模型的映射关系

定义中基本信度分配值 $m$ 是人根据证据做出的对命题看法，获得 $m$ 值后，就可以对不同证据源上的证据进行合成了。经典的证据组合公式要求所有参与合成的证据具有相同的重要程度，但在普适计算环境下，不同证据源得到的证据对目标的重要性与可靠性可能会不同，因此在证据合成前需要对证据的基本可信度进行加权修正<sup>[10,11]</sup>，以反映证据间不同的重要性与可靠性，修正信度公式如下：

$$m'_i(A_i) = \begin{cases} Bel(E)m_i(A_i), & A_i \neq \Theta \\ 1 - \sum m'_i(A_i), & A_i = \Theta \end{cases} \quad (2)$$

针对不同的上下文应用，获取Bel方法不同，但Bel的选择一般要遵循如下的规则：

(1)证据的重要性、可靠性越高，则对应的修正系数越大，即给重要的证据分配较大的权重。

(2)证据冲突越高，可靠性越低，则对应的修正系数越小，即要对它进行较大的修正，将冲突的信度部分更多地传给未知领域。

文献[12]提出了加权组合规则进行证据融合，解决了证据间高度冲突时证据理论的不足，但是该方法的推理模型不具有自适应性，因此本文结合修正

信度公式，改进了组合规则，公式如下：

$$m(A) = m_1 \oplus m_2(A) = \frac{\sum_{A_i \cap B_j = A} \omega_1 m'_1(A_i) \omega_2 m'_2(B_j)}{1 - \sum_{A_i \cap B_j = \phi} \omega_1 m'_1(A_i) \omega_2 m'_2(B_j)} \quad (3)$$

其中  $\omega_i = \frac{\sum_{j=1, j \neq i}^n (1 - d_{BPA}(m_i, m_j))}{\sum_{i=1}^q \sum_{j=1, j \neq i}^n (1 - d_{BPA}(m_i, m_j))}$ ， $d_{BPA}(m_i, m_j)$

为  $m_i, m_j$  的距离。

式(3)的特点：反映了证据间的不同的重要性和可靠性；解决了冲突证据组合中的“一票否决”问题及“鲁棒性”问题，充分利用冲突证据信息，避免了证据有效信息的损失。

### 3.3 推理算法流程

UCOMRADS实现算法流程如图3所示，具体步骤如下：

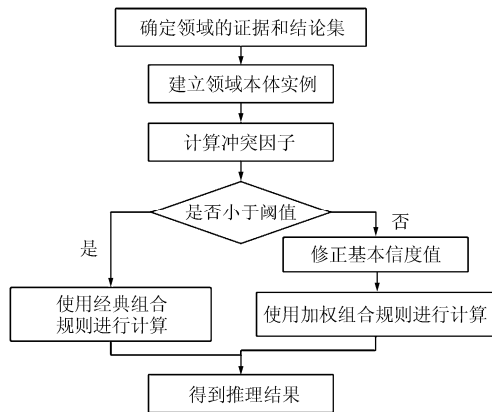


图3 不确定上下文推理流程算法

(1)对于领域本体模型，首先确定该领域内的结论和证据集，即Conclusions和Evidences，并列举出它们的全部子类ConclusionValue和EvidenceValue，依据本文第2节的上下文本体概念模型，建立相关领域不确定性上下文信息的本体实例。

(2)根据本文提出的映射关系，确定信任结构  $D$  中各个元素，依据所收集各证据的传感器类型、历史数据等信息，确定EvidenceValue的数据属性hasBel的取值和基本信度分配hasBPValue的取值。

(3)根据本节介绍的不确定性推理方法，进行推理。首先设定阈值  $\varepsilon$ ，根据式(1)计算冲突因子  $k$ ，若  $k > \varepsilon$ ，说明证据间高度冲突，首先依据式(2)计算修正后的信度函数值，然后根据式(3)计算出各个相应的ConclusionVulue的数据属性hasBel值；若

$k \leq \varepsilon$ ，则判定证据间冲突较小，依据式(1)计算出ConclusionValue的hasBel值。

(4)比较各个ConclusionValue的hasBel属性的值，最大的hasBel值即为最终的推理结果。

### 3.4 实验及分析

为了验证 UCOMRADS 方案的可行性、自适应性，本文实现了相应的原型系统 CASDSR。CASDSR 分为 3 层：模拟传感网络、上下文处理平台和上下文应用服务器。模拟传感网络是 CASDSR 的上下文信息来源；上下文处理平台是 CASDSR 的核心，是构建上下文感知应用的基础支撑软件，它独立于特定的上下文感知应用，分离了上下文信息的采集与感知应用的开发，向上对应用开发者提供统一的上下文信息使用接口，向下可以接入各类物理或逻辑传感器，主要功能模块包括采集器通信接口、智能体 Agent、推理引擎、应用接口、本体/规则库等；上下文应用服务器部署上下文应用，为用户提供个性化的服务。

不确定性知识的推理长期以来被认为是人工智能领域中一个重要研究问题，在许多专业领域中，如经济分析、医疗诊断、故障检测等，分析推理的过程中可利用的证据和知识往往是不确定的。故本文选择远程医疗诊断作为CASDSR的应用场景。在这个场景中，实时监测用户的生理数据，包括体温、血压、脉搏、呼吸、血糖、血氧饱和度、心率等，CASDSR通过这些上下文信息以及用户的病历信息进行诊断，实现用户的远程诊断或身体检查监护，以便得到救护。UCOMRADS需要对用户生理参数进行建模，然后进行不确定性推理，得出用户的疾病类型。下面分别介绍这两部分的实现过程，以及验证改进的组合规则的有效性，合理性。

**不确定上下文本体建模** 用户生理参数本体和用户偏好本体已经建立，现在只需要根据第2节提出的不确定本体概念模型，建立基于证据理论的不确定上下文模型，确定信任结构结论  $D = \langle \Theta, E, F, m \rangle$ 。在这个医疗应用场景中，取结论识别框架  $\Theta = \{\{Pneumonia\}, \{Heartfailure\}, Cardiomyopathy\}, \{\{Pneumonia\}, \{Heartfailure\}, \{Cardiomyopathy\}\}\}$ ，与  $\Theta$  相关的上下文特征信息即证据  $E = \{Temperature, BloodPress, Breath, Pulse\}$ ，然后确定每个上下文特征信息的状态空间：Temperature= $\{Highot, Lowhot, Nortem, Lowtem\}$ ，BloodPress= $\{Highpre, Lowpre, Norpre\}$ ，Breath= $\{Highbreath, Lowbreath, Norbreath\}$ ，Pulse= $\{Highpulse, Lowpulse, Norpulse\}$ 。最后根据医学知识，并咨询了医学领域专家意见，确定每个生理参数与高层推

理结果  $\Theta$  的映射关系  $F$ 。

**不确定上下文信息推理** 不确定推理过程就是使用组合规则进行证据融合的过程。本实验中, 信度分配表由医学领域专家给出, 还需要计算式(3)中的信度修正系数 Bel 值。为了使推理具有自适应性, 这里 Bel 的选择主要依据用户的历史数据, 具体确定方法如下:

(1)建立用户的病历表。表中保存用户名, 疾病类型(即推理结果), 推理结果出现的次数, 得到推理结果的时间等信息。每次执行推理时, 比较表中的推理结果, 更新次数和时间信息, 以便下次推理时计算 Bel。

(2)计算 Bel。设病历表中疾病类型为  $R_1, R_2, \dots, R_n$ , 相应的出现次数为  $t_1, t_2, \dots, t_n$ , 推理过程中使用的证据为  $e_1, e_2, \dots, e_m$ , 各证据的信度修正为  $Bel_1, Bel_2, \dots, Bel_m$ , 假设与  $R_i$  关联最大的证据为  $e_i$ 。找出  $t_i (0 < i \leq n)$  中的最大值  $t_i = t_{\max}$ , 则计算 Bel 的公式为

$$Bel_j = \frac{t_j}{t_{\max}}, \quad 0 < j \leq n, \quad Bel_i = 1 \quad (4)$$

确定了 Bel 后就可以利用本文改进的组合规则进行不确定性推理了, 并和文献[12]的方法进行比较, 验证本文提出的方法的有效性, 合理性。

由于冲突证据的存在, 使用经典的证据理论规则进行推理可能会得到不正确的结果, 而改进的良好组合规则应能减少这种情况。本文采集 10 组包含冲突证据的测试用例集, 容量分别为 1000, 2000, ..., 10000, 分别使用本文方法和文献[12]方法进行推理, 并统计出这 10 组测试用例的正确率, 实验结果如图 4 所示, (75%表示 1000 个测试用例中有 750 个推理结果是正确的,)表明随着测试用例容量的增加, 使用本文方法正确率也增加, 并且在大约 8000 个用例时趋于稳定, 而使用文献[12]方法的正确率和用例容量没有必然关系, 说明 UCOMRADS 能够根据历史数据和当前采集数据, 逐渐消除冲突证据对推理结果的负面影响, 不断调整自己的行为, 具有

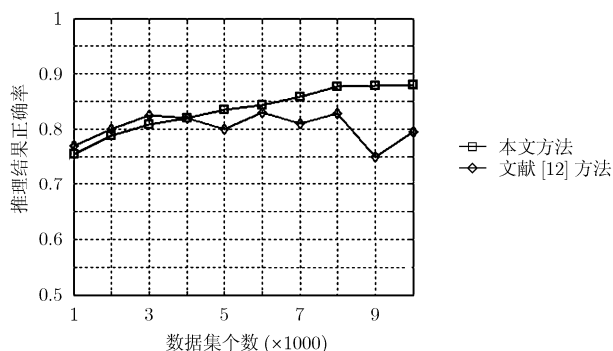


图 4 实验分析

自适应的能力。

## 4 总结

不确定性的知识表示与推理方法是研究和开发上下文感知系统的难点。本文首次将证据理论和本体建模结合起来, 提出了基于证据理论的上下文本体建模和推理方法, 并对证据组合规则进行了改进, 不仅解决了证据理论在高度证据冲突时的局限性, 而且使得该推理模型具有自适应性, 实现了不确定推理过程。实验表明基于证据理论的不确定上下文本体建模和推理具有可行性、自适应性。

## 参考文献

- [1] Ye J, McKeever S, and Coyle L, *et al.*. Resolving uncertainty in context integration and abstraction: context integration and abstraction [C]. Proceedings of the International Conference on Pervasive Services, New York, 2008: 131-140.
- [2] 乔秀全, 李晓峰, 廖建新. 基于贝叶斯网络的业务上下文认知模型构建方法[J]. 电子与信息学报, 2008, 30(2): 464-467. Qiao Xiu-quan, Li Xiao-feng, and Liao Jian-xin. Construction approach for Bayesian network-based service context recognition model [J]. *Journal of Electronics & Information Technology*, 2008, 30(2): 464-467.
- [3] Ko Kwang-Eun and Sim Kwee-Bo. Development of context aware system based on Bayesian network driven context reasoning method and ontology context modeling[C]. International Conference on Control, Automation and Systems, Seoul, Oct. 2008: 2309-2313.
- [4] Wu Huadong, Siegel M, and Stiefelhagen R, *et al.*. Sensor fusion using Dempster-Shafer theory[C]. IEEE Instrumentation and Measurement Technology Conference Anchorage, AK, USA, May 2002: 21-23.
- [5] Thomas O and Russomanno D J. Applying the semantic Web expert system shell to sensor fusion using Dempster-Shafer theory [C]. Proceedings of the Thirty-Seventh Southeastern Symposium, Tuskegee, March 20, 2005: 11-14.
- [6] 张德干, 徐光祐, 史元春, *et al.*. 面向普适计算的扩展的证据理论方法[J]. 计算机学报, 2004, 23(7): 918-927. Zhang De-gan, Xu Guang-you, and Shi Yuan-chun, *et al.*. Extended method of evidence theory for pervasive computing [J]. *Chinese Journal of Computers*, 2004, 23(7): 918-927.
- [7] Lu Wengxing, Liang Changyong, and Ding Yong. A method determining the object weights of experts based on evidence similarity in group decision-making[C]. Wireless Communications, Networking and Mobile Computing, Dalian, Oct. 2008: 1-4.
- [8] Liu Pei-zhi and Zhang Jian. A context-aware application infrastructure with reasoning mechanism based on Dempster-Shafer evidence theory[C]. Vehicular Technology

- Conference, Singapore, May 2008: 2834-2838.
- [9] 罗志增, 叶明. 用证据理论实现相关信息的融合[J]. 电子与信息学报, 2001, 23(10): 970-974.
- Luo Zhi-zeng and Ye Ming. Fusion of dependency information using Dempster-Shafer evidential reasoning [J]. *Journal of Electronics and Information Technology*, 2001, 23(10): 970-974.
- [10] Miao Y Z, Zhang H X, and Zhang J W, *et al.*. Improvement of the combination rules of the D-S evidence theory based on dealing with the evidence conflict[C]. 2008 IEEE International Conference on Information and Automation, Shanghai, June 2008: 331-336.
- [11] Sun Rui, Huang Hong-zhong, and Miao Qiang. Improved information fusion approach based on D-S evidence theory [J]. *Journal of Mechanical Science and Technology*, 2008, 22(12): 2417-2425.
- [12] Jousselme A, Grenier D, and Bosse E. A new distance between two bodies of evidence [J]. *Information Fusion*, 2001, 2(2): 91-101.
- 李艳娜: 女, 1985 年生, 硕士, 研究方向为网络管理与通信软件.
- 乔秀全: 男, 1978 年生, 讲师, 研究方向为融合业务网络及业务智能化支撑技术.
- 李晓峰: 女, 1950 年生, 教授, 研究方向为网络智能与通信软件.