

基于贝叶斯网络的业务上下文认知模型构建方法

乔秀全 李晓峰 廖建新

(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: 业务上下文信息的处理是未来泛在融合电信业务智能化的基础, 而不确定性是业务上下文信息不可避免的一个问题。该文利用贝叶斯网络理论, 提出一种支持不确定性推理的业务上下文认知模型的构建方法, 用来解决具有不确定性的业务上下文信息的融合和聚类, 并通过仿真实验验证了模型的合理性和可用性。

关键词: 电信业务智能化; 业务上下文认知; 贝叶斯网络; 不确定性推理; 上下文推理

中图分类号: TN915

文献标识码: A

文章编号: 1009-5896(2008)02-0464-04

Construction Approach for Bayesian Network-Based Service Context Recognition Model

Qiao Xiu-quan Li Xiao-feng Liao Jian-xin

(State Key Lab of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: The process of service context information is the foundation of future ubiquitous convergent telecommunication services intelligence. Uncertainty is an unavoidable problem of service context information process. In this article, based on Bayesian network theory, one construction approach for service context recognition model of supporting uncertainty reasoning is presented to resolve the convergence and classification of uncertain service context information. Furthermore, the rationality and usability of this model are verified by simulations.

Key words: Telecommunication service intelligence; Service context recognition, Bayesian network; Uncertainty reasoning; Context reasoning

1 引言

业务上下文指的是会影响到业务行为的相关信息。对于电信业务来说, 涉及的业务上下文信息有多种, 如接入网络的类型、通信协议、带宽、终端的能力、用户所处的物理环境(如位置, 天气情况)以及社会化人文环境(与周围的人以及各种活动的关系)等。业务上下文信息的处理是业务智能化、个性化的基石, 是业务动态自适应的关键技术之一。由于现实世界的复杂性、多变性, 导致存在大量的不确定性业务上下文信息, 如传感器采集的信息具有不精确性, 存在着一定的误差等。由此可以看出, 如何结合电信业务网络的特点, 真实地、完整地表示业务环境存在的信息, 以及对这些信息所形成的上下文环境进行认知, 是业务智能化所面临的一个关键问题。

目前已有许多关于上下文方面的研究^[1,2], 但主要局限于普适计算领域, 没有真正的与电信业务网络环境结合起来考虑^[3]。不确定性问题是普适(或泛在)计算环境下上下文感知应用不可避免的一个问题。目前, 已经引起了研究人员的极

大的兴趣^[4-7]。从现有的相关研究工作可以看出, 目前的大量研究工作主要集中在上下文信息的表示和推理方面, 体现在基于本体的上下文建模和推理, 以及将本体与概率相结合来研究不确定性上下文信息的表示和推理。在不确定性上下文信息处理方面, 现有的研究大都只作了不确定性信息的建模和表示方面定性的分析探讨, 缺乏系统有效地构建支持不确定性推理的上下文认知模型的方法, 更缺乏定量的实验分析和验证。另外, 还缺乏将上下文感知技术应用到电信业务领域, 以支持业务智能化和个性化的深入系统的研究。

本文在现有研究的基础上, 将上下文感知技术和电信业务网络相结合, 提出了一种基于贝叶斯网络的业务上下文认知模型的构建方法, 并将其应用到办公室环境下的智能呼叫业务的智能接入自适应处理方面, 增强了业务的智能性, 一定程度上弥补了确定性推理的局限性。

2 基于贝叶斯网络的业务上下文认知模型的构建方法

在业务的智能化机理中, 如何根据一些原始上下文特征信息来识别业务相关的当前高层上下文环境信息, 即用低层的、显式的上下文信息来推理高层隐含的上下文信息是上下文信息汇聚和推理的基本特征。可以利用上下文信息之间的

2006-07-05 收到, 2006-12-20 改回

国家自然科学基金(60672122), 国家杰出青年科学基金(60525110)和高等学校博士学科点专项科研基金(20070013026)资助课题

这种依赖或因果关系来构建贝叶斯网络。在利用贝叶斯网络来进行业务上下文的推理时, 由于存在多种不确定性因素, 如根据先验知识构造贝叶斯网络的主观性、训练样本集的局限性(如样本的容量、样本数据的缺失情况)等, 使得业务上下文认知模型的构建、推理、结果的反馈、模型的再学习是一个不断循环和完善的进化过程, 最终使得业务上下文认知模型能够尽可能准确地反映客观现实。本文利用此思想, 提出了一种基于贝叶斯网络的业务上下文认知模型的构建方法(如图 1 所示)。

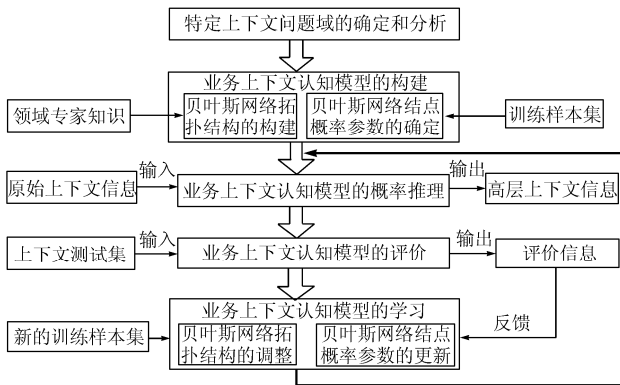


图 1 基于贝叶斯网络的业务上下文认知模型构建方法

在构建基于贝叶斯网络的业务上下文认知模型时, 主要分为以下几步。

2.1 特定上下文问题域的确定和分析

首先需要针对不同的业务需求和场景, 明确业务所涉及到的具体的上下文问题域, 即分析问题域的边界、求解目标和相关的前提条件等因素, 分析涉及到的上下文信息(或事件)之间的因果依赖关系和条件独立性关系, 并赋予明确的物理含义解释。为下一步的数学模型的抽象和建模奠定基础。

2.2 业务上下文认知模型的构建

根据贝叶斯网络的定义, 构建业务上下文认知模型主要包括两部分工作, 即贝叶斯网络的拓扑结构和结点的概率分布。

(1)业务上下文贝叶斯网络拓扑结构的构建 在确定上下文认知的贝叶斯网络结构时, 先确定上下文认知模型有关的变量及其解释。首先, 确定上下文认知的目标变量(C), 明确其代表的实际物理含义, 并确定其可能的取值 (C_i)和状态空间数 $N (i=1, \dots, N)$, 结点的每个取值对应一个具体的高层业务上下文信息; 其次, 确定与该认知目标相关的上下文特征向量 $F(F_1, F_2, \dots, F_M)$, F_i 表示不同的上下文特征变量, 并分析这些上下文特征之间以及与目标变量之间的因果依赖关系, 进行条件独立性分析, 形成贝叶斯网络的有向无环图结构; 最后, 分别确定每个上下文特征变量 F_i 的取值和状态空间数, 每个取值分别对应一个原始的、直接感知到的业务上下文信息。通过这 3 步, 就可以确定贝叶斯网络的拓

扑结构。然后获取相应的上下文训练数据集, 进行网络结点的概率分布的学习。

(2)业务上下文贝叶斯网络结点概率分布的确定 如果缺乏训练样本数据集, 则需要利用领域专家的先验知识来指配结点的初始概率分布。本文主要针对在具有训练样本数据集情况下的结点概率参数学习问题, 主要包括两种情况, 一种是具有完整的样本数据集, 即所有的网络结点变量的值都是可以观察到的; 另一种就是由于某种原因, 导致训练样本数据集不完整, 存在数据缺失的情况, 即有的网络结点变量被隐藏起来, 无法获取其观察值。基于贝叶斯网络的样本训练集可以表示为 $D = \{D_1, D_2, \dots, D_m\}$, 其中 $D_i = \{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\}$ 表示所有变量构成的向量 $X = (X_1, X_2, \dots, X_n)$ 的一个实例。如果一个实例中所有的变量取值确定, 称为一个完整实例, 否则为不完整实例。由完整实例构成的样本训练集称为完整训练数据集, 否则称为不完整训练数据集。

(a)利用完整样本数据集进行概率参数学习的方法 在网络结构已知的情况下, 利用完整样本训练集学习概率参数一般采用最大似然估计(MLE)的方法, 假设所有结点变量的联合概率密度为 $P(X)$, θ 是结点概率参数组 $(\theta_1, \theta_2, \dots, \theta_n)$ 的向量, 则在已知样本数据集 D 的情况下, 参数向量 θ 的似然函数为 $L(\theta | D) = P(D | \theta) = \prod_{i=1}^m P(D_i | \theta)$, 似然函数可以看

作是在 D 固定的情况下关于参数 θ 的函数。最大似然估计的方法就是寻找合适的 θ 使得似然函数 L 最大, 为了计算方便, 一般通过求使似然函数 L 的对数 $\text{Log}(L(\theta | D))$ 函数取最大值的 θ , 取对数将概率相乘转化为相加, 从而简化计算。

(b)利用缺失样本数据集进行概率参数学习的方法 在网络结构已知, 训练样本数据存在缺失的情况下, 一般采用期望最大化(EM)算法来进行参数的估计, 主要包括两步, 即 E 步骤和 M 步骤。E 步骤先求每个样本实例缺失变量的期望, 将该期望值当作观察到的值来使用, 将缺失的样本训练数据补全, 从而形成一个完整的训练样本集; M 步骤就是在完整训练样本集的情况下, 利用最大似然估计的方法来寻求使似然函数的期望最大化的新参数值。然后在新参数的基础上, 进行新一轮的 E 步骤和 M 步骤, 该过程不停迭代, 直到收敛为止。假设 θ 是结点概率参数组 $(\theta_1, \theta_2, \dots, \theta_n)$ 的向量, D 表示不完整的训练样本数据集, $Z = (D, G)$ 表示一个完整的样本数据集, 其中 G 是对 D 补全的数据集, 所有结点变量的联合概率密度为 $P(X)$, 则有似然函数 $L(\theta | Z) = L(\theta | D, G) = P(D, G | \theta)$, EM 算法就是要寻找合适的 θ , 使得 $E(\text{Log}(P(D, G | \theta) | D, \theta^{i-1}))$ 的值最大, 其中 θ^{i-1} 是用来计算期望的当前的结点概率参数估计。

2.3 业务上下文认知模型的概率推理

构建好基于贝叶斯网络的业务上下文认知模型后, 就可以初步地应用到实际系统中, 对从真实上下文环境中感知、

收集来的原始上下文信息进行聚类或因果分析,推断出高层的上下文信息。基于贝叶斯网络的上下文认知模型的推理就是利用给定的上下文变量信息(也叫证据)来计算其它变量的概率信息的过程。假定直接感知到的原始上下文信息用向量 $\mathbf{F}(F_1, F_2, \dots, F_M)$ 来表示,高层上下文信息用类别向量 $\mathbf{C} = (C_1, C_2, \dots, C_n)$ 来表示,则上下文认知模型的推理就是计算 $P(C_i | \mathbf{F})$,从而选出可能性最大的高层上下文信息 C_i 最为推理结果。本文采用了选择后验概率最大的评判方法,即 $P(C_i | \mathbf{F}) \geq P(C_j | \mathbf{F})$,此时取判别函数 $r_i(\mathbf{F}) = P(C_i | \mathbf{F})$ 。理论证明,这种方法能够保证分类误差最小。业务的智能决策部分利用得到的高层业务上下文信息,就可以制定相应的业务适配策略。

2.4 业务上下文认知模型的评价

构建的上下文认知模型是否满足业务的实际要求,需要对模型进行进一步的评估。选取一定容量的测试样本数据集,对已建好的业务上下文认知模型的推理精确度进行检验和评价。假设测试数据集为 $D = (F_1, F_2, \dots, F_M, C)$,其中 F_i 表示感知到的上下文信息, C 表示需要认知的高层上下文信息。针对测试集 D 的每个样本实例 $d_i(f_{1i}, f_{2i}, \dots, f_{Mi}, c_i)$,将感知到的原始上下文信息 $(f_{1i}, f_{2i}, \dots, f_{Mi})$ 作为上下文认知模型的输入证据,判断概率推理的结果是否符合实际的样本值 c_i ,即判断 $P(c_i | f_{1i}, f_{2i}, \dots, f_{Mi}) \geq P(c_j | f_{1i}, f_{2i}, \dots, f_{Mi})$,其中 $i \neq j$,且 $0 < i, j \leq n$, n 为高层上下文信息变量 C 的状态空间数。对推理结果进行统计,从而得出该模型针对该测试集的精确度。如果不符合要求,则进入下一步,根据反馈信息重新对业务上下文认知模型进行学习。

2.5 业务上下文认知模型的学习更新

上下文认知学习模型利用评价阶段的反馈信息和新产生的大量的样本数据来对上下文认知的贝叶斯网络模型进行重新学习,可能包括调整贝叶斯网络的拓扑结构或者更新结点变量的概率分布。一方面,可以通过样本数据来学习网络结构,然后结合先验知识来对网络拓扑结构进行调整;另一方面,可以增大样本容量来提高结点变量的概率分布的学习精度。从而保证支持上下文认知的贝叶斯网络模型能够在同上下文环境进行交互的过程中进行不断的学习和更新,达到不断的完善和进化。

3 实验与分析

本文以办公室当前的活动情况的上下文信息作为业务适配的依据,对呼入电话进行智能化处理。如当办公室主人正在与其他人交谈,则为了避免打扰谈话,呼入的电话可以转接到语音信箱或将该呼叫放入等待列表,当谈话结束后由应用发起呼叫将两人的电话接通。

(1)业务上下文贝叶斯网络拓扑结构的构建 将办公室的活动作为贝叶斯网络的目标查询结点 C ,它有4个状态 C_i , $i=1, \dots, 4$,分别是 {Idle, Conversation, Phoning,

Working}。本文主要选取了办公室人数(NumberOfPeople)、手机状态(MobilePhoneStatus)、办公室固定电话状态(FixedPhoneStatus)、噪声水平(NoiseLevel)这4种上下文信息来进行办公室活动的推理。每个上下文特征信息的状态空间为: NumberOfPeople {=0, =1, >1}, MobilePhoneStatus {Off, Idle, Busy}, FixedPhoneStatus {Idle, Busy}, NoiseLevel {High, Low}。每个上下文信息特征变量的取值对应着一个上下文信息,根据这些上下文信息间的因果关系及条件独立性属性,构建了如图2所示的认知办公室活动状态的贝叶斯网络拓扑结构图。

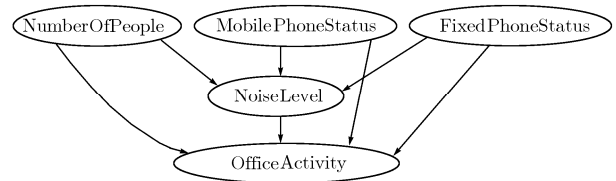


图2 认知办公室活动状态的贝叶斯网络拓扑结构

(2)业务上下文贝叶斯网络的结点概率参数学习 在构建了贝叶斯网络模型的有向无环图后,需要给每个结点指派相应的概率分布。由于先验概率参数可以通过样本数据的学习来获取,因此实验先采用随机的方式指派结点的初始概率分布,然后通过样本数据的学习对贝叶斯网络进行训练,更新结点的概率参数,使其符合实际的办公室上下文环境。

(a)完整样本数据集的概率参数学习 图3反映了在办公室有1人、手机状态为空闲、办公室座机忙以及室内噪声高的情况下,不同容量大小的完整样本空间对办公室4种状态的概率估计的影响变化图。可以看出,训练样本容量的大小对贝叶斯网络的概率参数的准确度有不同程度的影响。样本容量越大,则学习的概率参数应该更加接近于真实情况。从图3(a)可以看出,接近80%的可能是处于通话状态,在样本容量超过9万之后,通话状态概率值的变化就趋于稳定。图3(b)反映了其它3种状态的概率都很小。

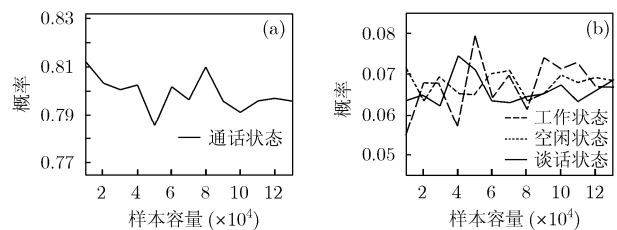


图3 不同容量的样本空间下办公室各种活动状态的概率分布变化图

在构建好用来推理办公室活动的业务上下文贝叶斯网络模型后,就可以利用该网络模型来对实际感知到的上下文特征信息进行高层推理。本文采用了贝叶斯网络的联合树算法来进行精确推理。在实验中,利用上下文感知模拟环境按

照用户的日常习惯采集了容量为 10000 的测试用例集, 每个测试用例 $case_i$ 的结构为 {NumberOfPeople, MobilePhoneStatus, FixedPhoneStatus, NoiseLevel, OfficeActivity}。因此可以将每个测试用例的前 4 个参数的取值作为证据, 用上下文认知的贝叶斯网络模型来计算 OfficeActivity 的最可能取值, 然后和测试用例中实际的 OfficeActivity 的取值进行比较, 从而对上下文认知模型的推理准确率进行统计。实验结果表明, 在作为证据的上下文信息完整的情况下, 基于贝叶斯网络的上下文认知模型的准确率达到了 82.23%。

另外, 也对有缺失上下文信息的情况下, 进行高层信息的推理进行了用例测试。在实验模拟中, 将测试用例集中作为推理证据的前 4 个参数的部分数据抹掉, 形成不完整测试用例集, 然后推理办公室活动。实验结果表明, 此时基于贝叶斯网络的上下文认知模型的准确率达到了 65.4%。

(b) 缺失样本数据集的概率参数学习 考虑到现实环境中, 由于各种原因(如传感器故障)可能导致存在的样本数据的缺失情况, 在实验中利用容量为 10000 和 20000 的样本数据, 分别构造了不同缺失程度(10%到 70%)的新的不完整样本集, 采用期望最大化(EM)算法对贝叶斯网络结点概率参数进行了学习。对于办公室有 1 人、手机状态为空闲、办公室座机忙以及室内噪声高的情况下, 办公室处于通话状态的情况进行了分析, 图 4(a) 比较了样本容量分别为 10000 和 20000 的情况下, 不同缺失程度下学习到的概率参数情况。可以看出, 随着缺失数据量的增大, 概率参数的学习精度在大幅下降; 而在同等缺失条件下, 增大样本容量空间可以提高概率参数学习的精度。图 4(b) 以 20000 容量的样本数据为例, 说明了不同数据缺失情况下 EM 算法在门限为 0.001 的情况下收敛情况, 可以看出, 样本数据缺失程度越大, 需要迭代的次数越多, 收敛速度越慢。

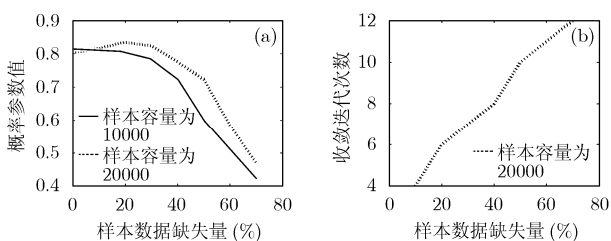


图 4 样本数据缺失情况下 EM 算法的概率参数学习结果

4 结束语

对业务上下文信息的处理, 是业务智能化的关键技术之一。本文基于贝叶斯网络理论, 提出了一种支持不确定性推理的业务上下文认知模型的构建方法, 从而可以与现有的上下文信息建模和推理技术相结合, 有效地支持业务上下文的认知。本文将上下文感知技术和概率统计理论应用到电信业务领域, 在探索业务智能化支撑理论和技术方面进行了积极尝试。

参考文献

- [1] Dey A K and Abowd G D. Towards a better understanding of context and context awareness. Technical Report GIT-GVU-99-22, College of Computing, Georgia Institute of Technology, 1999.
- [2] Wang X H, Dong J S, and Chin C Y, *et al.* Semantic space: an infrastructure for smart spaces. *IEEE Pervasive Computing*, 2004, 3(3): 32-39.
- [3] Wang Xiaohang, Gu Tao, and Zhang Daqing, *et al.* Ontology based context modeling and reasoning using OWL. Workshop on Context Modeling and Reasoning (CoMoRea) at IEEE International Conference on Pervasive Computing and Communication (PerCom'04), Orlando, Florida. March 14, 2004: 18-22.
- [4] Satyanarayanan M. Coping with uncertainty. *IEEE Pervasive Computing*, 2003, 2(3): 2.
- [5] Abdelsalam W and Ebrahim Y. Managing uncertainty: modeling users in location-tracking applications. *IEEE Pervasive Computing*, 2004, 3(3): 60-65.
- [6] Ranganathan A, Al-Muhtadi J, and Campbell R H. Reasoning about uncertain contexts in pervasive computing environments. *IEEE Pervasive Computing*, 2004, 3(2): 62-70.
- [7] Gu T, Pung H K, and Zhang D Q. A Bayesian approach for dealing with uncertain contexts. http://www.comp.nus.edu.sg/~gutao/gutao_NUS/pervasive2004_gutao.PDF, April 2004.

乔秀全: 男, 1978 年生, 博士, 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。

廖建新: 男, 1965 年生, 教授, 研究方向为移动智能网、下一代网络、3G 移动通信。