DOI:10.3969/j.issn.1673-4785.201407009
网络出版地址:http://www.cnki.net/kcms/detail/23.1538.TP.20150302.1106.008.html

基于语义分层的行为推理框架

聂慧饶,陶霖密
(清华大学 计算机科学与技术系,北京 100084)

摘 要:人类行为理解是实现“人本计算”模式的基础,其本质在于获取行为的语义,即由动作特征推导人体的行为,需要跨越两者之间的语义鸿沟;为此提出了环境上下文进行隐式建模的方法,并基于此提出了语义分层的行为推理框架,该框架使用了从模糊语义到确定语义的渐近式推理。根据知识将特征合理地分为多个层次,系统则根据当前状态去提取所需要的特征,推理当前可能的候选行为集;并由该候选行为集指导处理模块,更新特征集并进行新一轮的推理,反复迭代至推理完成。应用提出的环境建模方法和渐近推理框架可以有效地实现行为理解。使用隐式环境方法可以提高行为理解的准确率;渐近式推理框架可以避免传统推理方法无差别地提取所有特征,从而提升了推理效率。

关键词:行为理解;特征行为关系;环境上下文;语义分层;分层推理框架
中图分类号: TP301.6 **文献标志码**:A **文章编号**:1673-4785(2015)02-0178-09

中文引用格式:聂慧饶,陶霖密. 基于语义分层的行为推理框架[J]. 智能系统学报, 2015, 10(2): 178-186.
英文引用格式:NIE Huirao, TAO Linmi. Inference framework for activity recognition based on multiple semantic layers[J]. CAAI Transactions on Intelligent Systems, 2015, 10(2): 178-186.

Inference framework for activity recognition based on multiple semantic layers

NIE Huirao, TAO Linmi
(Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China)

Abstract: Human activity recognition is the core of the implementation of human-centered computing (HCC), whose nature is to acquire activities′ semanteme. The basic problem is the semantic gap between observable actions and human activities. They should be bridged by environment context based inference. In this paper, a method is proposed to model the environment context implicitly. Further, a novel semanteme multilayered activity inference framework was presented, which divided the inferring process into 2 stages. One stage used to acquire fuzzy semanteme and another one to acquire accurate semanteme. The feature set was divided into different subsets according to knowledge. The system extracts the corresponding features according to the current state and obtains the possible set of candidate activities that can instruct the system to update the current feature set. Update the features set and infer it, the process continues until the inference is completed. The modeling method and progressive inference framework proposed could handle the activity-recognition problem well. Implicitly modeling the environment context could improve the accuracy of activity recognition. The progressive framework can improve the efficiency by avoiding extracting all features indistinguishably, whose validity was proven in the data set.

Keywords: activity recognition; feature activity relation; environment context; semantic layer; multilayer inference framework

收稿日期:2014-07-04. **网络出版日期**:2015-03-02.
基金项目:国家“863”计划资助项目(2012AA011602);国家自然科学基金资助项目(61272232).
通信作者:聂慧饶. E-mail:sangoblin@yeah.net.

Pantic 等[1]提出了“人本计算”(human-centered computing,HCC)的概念;这种模式被认为是未来的计

算模式,在该模式当中,计算被隐藏在居住空间的后台,而其计算结果则在日常生活当中与人交织在一起。与过去“以计算机为中心”的计算模式相比,HCC 使用了更接近人类交互方式的方法,如理解人类的行为和情感等,从而取代传统的键盘和鼠标输入;人类得以从过去僵化的输入环境当中解放出来,而使用更加贴近其天性的自然方式与计算机进行交互。

HCC 的研究重点在于使计算设备与传感设备进行协同工作以便主动感知场景中的用户信息,分析用户需求并完成相关任务[2]。因此,利用计算设备和传感器协同工作以理解人类的行为是 HCC 的核心组成部分。针对传统行为理解系统无差别提取场景当中所有特征的弊病,本文对行为所搭载的语义进行分层,并相应地对场景中的特征进行了分类,从而提出了一个由粗至精的逐步获取行为语义的推理框架。

1　研究现状

行为理解是计算机视觉领域的传统问题[3],其推理方式可以大致分为基于规则的推理和基于学习的推理[4]。基于规则是指研究者根据自己对行为逻辑的认识,并利用逻辑推理的方法对行为进行理解。该方法通常包含以下步骤:1)将所有可能的需要理解的行为囊括到模型库当中,并利用逻辑形式对这些行为进行定义和描述;2)整理所获得的传感信息,并将其转换为逻辑术语和公式;3)根据上一步当中的术语和公式,进行包括演绎、归纳和推断等的逻辑推理,以便于根据所观察的信息寻找最匹配的行为或者行为集(模型库的子集)[5-7]。

基于学习的推理则又可进一步细分为无监督学习和有监督学习。其中无监督的学习指的是从未进行人工标注的数据当中直接建立起模型对行为进行判别,其通用原则是根据系统当前的状态并结合对系统的观察对系统的状态进行随时更新,模型中每个动作可能发生的概率均是由人工进行赋予的;无监督学习的过程通常有:1)采集原始传感数据(未被标注)[8];2)处理未标注的数据并将其转换成相应特征;3)采用聚类等手段建立起判别模型[9-10]。与无监督的学习相比,有监督的学习必须基于已经标注的数据(通常是人工标注),而后根据数据和行为集建立起合理的推理模型,并通过标注数据训练出模型的参数。当前通过有监督的学习得到推理模型参数的方法是最为常见的,并且研究者们也在此方面总结了很多有效的算法和模型,如隐马尔可夫模型[11-13]、贝叶斯网络[14]、条件随机场[15]、最近邻法[16]等。

但是当前的推理方法当中均未考虑对行为的语义进行分层,并根据需要从环境当中提取特征,而是尽可能多地从环境当中提取特征后进行行为推理。

2　环境上下文模型

对行为进行推理时,若能引入人所处的环境上下文,则可以提高推理结果的精度。不少研究者在开展他们的工作时也引入了环境上下文的概念,但是他们通常是根据本体论将环境上下文显式地建立在了模型当中[17-18],该方法的缺点是:1)模型的可扩展性差,一旦环境有更改,需要重新建立一套模型;2)难以将时间上下文同时引入到模型当中。本文当中为了使环境上下文便于计算,未将其作为显式的模型节点,而是将其作为隐式的观测特征用于辅助行为的推理。

2.1　特征的属性

当前的相关工作大都采用了分层模型来表示行为,并将行为定义成了语义的携带者[19-21];行为通常都是为了满足用户需求而发生的一系列动作。伴随着行为的发生,通常可以观测到与该行为相关的特征。而行为理解需要处理的问题就是根据所观测的特征还原出用户的行为。

特征作为样本的表现形式,可以用于将一个样本与其他样本进行区分,例如,发生吃饭行为时,手中的餐具可以作为用来表征该行为的重要特征。因此,特征可以视作对样本的某种属性的观测。理论上,若能获取正确表达某个样本的完整的特征集,则可以以极高的置信度识别该样本。但是在基于视觉的处理方法当中,系统可以从视频图像中提取出大量不同的特征,如颜色直方图、SIFT 特征、HOG 特征等,而且基于视觉特征进行分类得到的结果通常具有不确定性。因此,当样本集的规模变得很大时,即面临着组合爆炸的问题,特征的规模会增长得比样本更快,很难一次性将视频中所有的特征悉数提取出来。

在本文当中,根据特征是否被行为集中的所有元素共享将其分为:公有特征和私有特征,其中公有特征属于某个行为集中的所有行为,即所有行为发生时该类特征都可被观测(但是特征值不同);私有特征则是某个行为所特有,通常可以用于证明或者证伪该行为是否发生。显然公有特征集和私有特征

集的选取依赖于特定的应用场景,并且可以根据应用需要对行为集进行多层次的分层,从而实现推理层次更加丰富的推理过程。

2.2 可计算的环境上下文

显然所有的行为发生都伴随着环境上下文,因此环境上下文应该属于公有特征。环境上下文是一个很抽象的概念,为了能将其予以形式化的表达,需要考虑用户在室内的交互方式以及交互对象;通常用户在室内的交互对象主要是各色家电以及家具,而用户的交互方式又由他当前的交互对象所决定,例如,用户处在卧室当中,则其可能在与床进行交互,交互方式则是用户躺在床上。而这些家具或者家电与特定行为的发生具有很强的关联,比如餐桌附近吃饭发生的概率很高;于是可以将它们的中心作为某些行为发生的概率中心,而随着人体逐渐的远离该行为发生的概率会逐渐衰减。

因此,可以将家具和家电等潜在交互对象的位置予以标定(如图 1 所示,图中的圆形和方形分别代表了室内的餐桌、冰箱等交互对象),并结合人体的当前位置作为观测的特征;在实现时,通常使用人体位置与各个交互对象间的坐标差值(或基于差值的非线性变换)作为观测特征,而不是使用人体到交互对象中心的欧氏距离,以考虑交互对象的形状对于行为发生的概率衰减的影响。

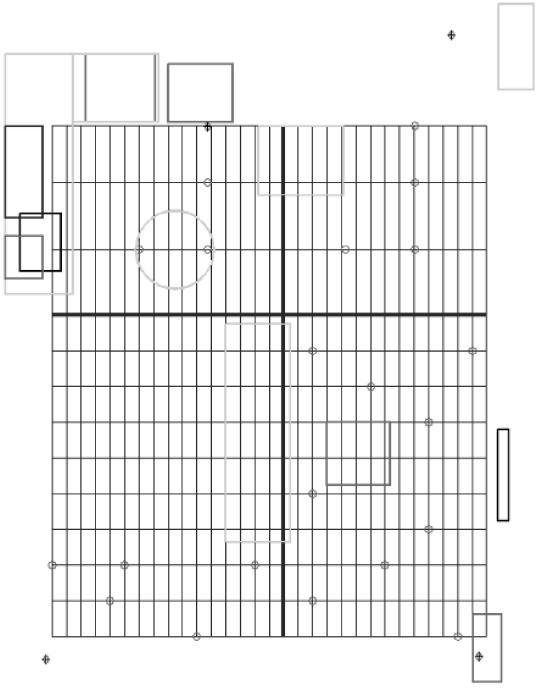


图 1 室内家具布置标定示意图

Fig.1 Calibration of the indoor layout of the furniture

3 分层的语义推理及实现

3.1 分层的语义推理

前文当中根据特征是否被行为集中的所有元素所共享将其分成了公有特征和私有特征 2 类。行为是具确定语义的,但是在观测到属于某个行为的所有特征前,尤其若其私有特征尚未被观测,则该行为的确定语义将无法被推断,其所携带的语义将变得模糊不清,从而该行为的确定语义将退化成为模糊语义。图 2 中所示,用户分别发生了 2 个行为,即喝水和喝饮料;这 2 个行为的公有特征即为手中持有物品,且在手部在向面部运动,而喝水的私有特征则是手中物品为水杯,喝饮料的私有特征为手中物品为饮料。显然这 2 个行为的公有特征是几乎一致的,区分它们的关键因素在于这 2 个行为不同的私有特征;但是公有特征的观测可以排除用户发生看电视等其他行为。

图 2(a)对喝水和喝饮料的私有特征进行了模糊化处理(即不再观测这 2 个行为的私有特征)后,喝水和喝饮料均退化成为语义模糊不清的动作,该动作表明人手中有物品且在向面部运动。从中可以看出,行为的公有特征即表达了行为的模糊语义,而辅以相应的私有特征后行为的语义才能被确定。

(a)原图

(b)观测私有特征

图 2 由粗至精的推理过程

Fig.2 Inference process from fuzzy semantic logic to definite semantic logic

因此,提出了一个从模糊语义逐渐到确定语义的推理框架,即首先根据观察到的公有特征筛选出符合当前模糊语义的候选行为集合${}_cA$,然后根据${}_cA$中的成员做证据广播,即去观测该成员的相应私有特征,并最终得到当前用户的行为或者行为集 A_{curr}

(用户可以同时发生多个行为)。事实上人的推理过程也并非一次完成的,人们总是会根据当前观测的特征对即将发生的行为作出初步的判断,而后根据初步判断的结果去寻找可以证明或者证伪初步判断的新的特征。例如,甲向乙伸出手时,乙初步判断甲想要同乙握手或者攻击乙,此时乙开始寻找额外的特征,若甲的手向乙的运动而去,则甲想要同乙握手,反之则是要攻击乙。

假定系统中共有 N 个行为需要识别,将该原始行为集记作 ${}_oA=\{a_1,a_2,\cdots,a_N\}$;系统可以获得的总特征数为 M 个,总特征集记 ${}_oF=\{f_1,f_2,\cdots,f_M\}$。这些特征包含了公有特征集以及私有特征集中的元素,其中 ${}_pF_k^j=\{{}_pf_{k1}^j,{}_pf_{k2}^j,\cdots,{}_pf_{k\,{}_pM_k^j}^j\}$ $(j=1,2,\cdots,J;k=1,2,\cdots,K_j)$ 为公有特征集,含有 ${}_pM_k^j$ 个元素,而 J 为推理的总次数,K_j 为第 j 次推理时候选行为集合 ${}_cA^j$ 的总个数;私有特征集针对 ${}_oA$ 中的每个元素进行定义,因此可以得到 N 个行为的私有特征集,${}_sF_i=\{{}_sf_{i1},{}_sf_{i2},\cdots,{}_sf_{i\,{}_sM_i}\}$ $(i=1,2,\cdots,N)$,对于第 i 个私有特征集,其含有 ${}_sM_i$ 个元素。则有

$$\begin{cases}\{\cup_{i=1}^{K_{J-1}}{}_pF_i^{J-1}\}\cup\{\cup_{i=1}^{N}{}_sF_i\}={}_oF\\ \{\cup_{i=1}^{K_{J-1}}{}_pF_i^{J-1}\}\cap\{\cup_{i=1}^{N}{}_sF_i\}=\varnothing\end{cases}$$

推理时则如前所述,第 j 次推理时可以先依据当前的公有特征集 ${}_pF_{\text{curr}}^j$ $(\text{curr}\in\{1,2,\cdots,K_j\})$ 从 ${}_cA^{j-1}$ 中得到模糊语义满足观测的候选集 ${}_cA^j$,${}_cA^j=\{{}_ca_1^j,{}_ca_2^j,\cdots,{}_ca_{{}_cN_j}^j\}$,${}_cN_j$ 是 ${}_cA^j$ 的元素的个数且

$$\begin{cases}{}_cA^0={}_oA,{}_cA^J=A_{\text{curr}}\\ {}_cA^j\subseteq{}_cA^{j-1},j=1,2,\ldots,J\end{cases}$$

通常 ${}_cN_J\ll N$。在第 J 次推理时,遍历 ${}_cA^{J-1}$ 的成员 ${}_ca_i^{J-1}$ $(i=1,2,\cdots,{}_cN^{J-1})$,根据对 ${}_ca_i^{J-1}$ 的私有特征集 ${}_sF_i$ 中的成员进行观测,可以得出 A_{curr}。该过程如图 3 所示。需要注意的是,该推理框架可以通过不同的推理方法予以实现,后文当中分别使用了逻辑回归和 HMM 对框架进行了实现。

3.2　基于单帧的推理方法

基于单帧的推理方法,即在推理时只使用当前视频帧所观测的特征进行行为理解,其优点在于推理方式相对简单,计算量较小,可以快速地完成,而其缺点在于不使用时间上下文,从而对于噪声的抗性较差。基于单帧的推理方法有逻辑回归、支持向量机以及决策树等。本文实现时使用了逻辑回归模型作为实现单帧推理的方法,可以方便地得到对齐到(0,1)的概率值,也便于处理同一时刻下多个不同行为的发生。

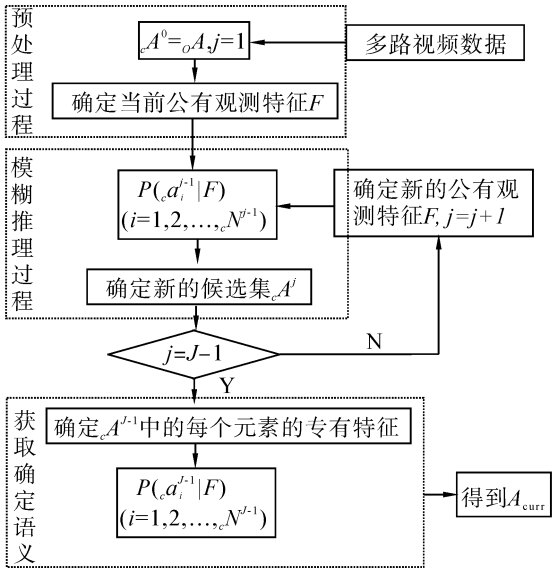


图 3　分层推理流程图

Fig.3　Flow chart of multilayer inference

进行分层次推理时,需要训练出多组逻辑回归模型以适应不同的推理层次。令 ${}_p\boldsymbol{X}_k^j=[{}_px_{k1}^j\quad {}_px_{k2}^j\quad\cdots\quad {}_px_{k\,{}_pM_k^j}^j]$ $(k\in\{1,2,\cdots,K_j\})$ 为第 j 次推理时观测到的公有特征向量,即 ${}_px_{ki}^j$ $(i=1,2,\cdots,{}_pM_k^j)$ 对应于 ${}_pf_{ki}^j$ 的观测值,第 j 次推理时的特征对行为的贡献度为 ${}_p\boldsymbol{W}_{zk}^j=[{}_pw_{z,k1}^j\quad {}_pw_{z,k2}^j\quad\cdots\quad {}_pw_{z,k\,{}_pM_k^j}^j]$ $(z=1,2,\cdots,{}_cN_{j-1})$,其中 ${}_pw_{z,i}^j$ 即表示第 j 次推理时 ${}_pf_{ki}^j$ 对 a_n 的影响。则第 1 次到第 $J-1$ 次推理时用到的逻辑回归模型为

$$\begin{cases}P({}_ca={}_ca_z\mid{}_p\boldsymbol{X}_k^j)=\\ \qquad\dfrac{\exp({}_p\boldsymbol{W}_{zk}^j\cdot{}_p\boldsymbol{X}_k^j)}{1+\sum_{i=1}^{N-1}\exp({}_p\boldsymbol{W}_{ik}^j\cdot{}_p\boldsymbol{X}_k^j)},z=1,2,\cdots,{}_cN_{j-1}-1\\ P({}_ca={}_ca_z\mid{}_p\boldsymbol{X}_k^j)=\\ \qquad\dfrac{1}{1+\sum_{i=1}^{N-1}\exp({}_p\boldsymbol{W}_{ik}^j\cdot{}_p\boldsymbol{X}_k^j)},z={}_cN_{j-1}\end{cases}$$

对于第 j 层逻辑回归模型,设定阈值 Threshold^j,若 $P({}_ca={}_ca_z\mid{}_p\boldsymbol{X}_k^j)>\text{Threshold}^j$,则 a_z 会被添加至 ${}_cA^j$。对于第 J 次推理,需要判断 ${}_cA^{J-1}$ 中的每个成员是否属于 A_{curr},因此需要对于所有的行为都训练一个二项逻辑回归模型,即根据特征判断该行为发生或者没有发生,对于行为 ${}_ca_m^J$ $(m\in\{1,2,\cdots,{}_cN_{j-1}\})$,第 J 次推理时用到的特征向量 $\boldsymbol{X}_m^J$ 为

${}_p\boldsymbol{X}_m^{J-1}$ 与 ${}_s\boldsymbol{X}_m$ 的并集，其中 ${}_s\boldsymbol{X}_m$ 为 ${}_ca_m^J$ 的私有特征对应的观测值，故对于 ${}_ca_m^J$ 有

$$\begin{cases} P({}_ca = {}_ca_m^J \mid \boldsymbol{X}_m^J) = \dfrac{\exp(\boldsymbol{W}_m^J \cdot \boldsymbol{X}_m^J)}{1 + \exp(\boldsymbol{W}_m^J \cdot \boldsymbol{X}_m^J)} \\ P({}_ca \neq {}_ca_m^J \mid \boldsymbol{X}_m^J) = \dfrac{1}{1 + \exp(\boldsymbol{W}_m^J \cdot \boldsymbol{X}_m^J)} \end{cases}$$

此时只需要将 $P({}_ca = {}_ca_m^J \mid \boldsymbol{X}_m^J)$ 与 Threshold^j 进行对比，即可判断 ${}_ca_m^J$ 是否属于 A_{curr}。需要注意的是，所有的行为的二项逻辑回归的值的和并没有归一化，所以在选取 Threshold^j 的值时，需要根据当前计算得到的各个 $P({}_ca = {}_ca_m^J \mid \boldsymbol{X}_m^J)$ 的值予以动态设置。

3.3 基于时间序列的推理方法

基于时间序列的推理方法主要有隐马尔可夫模型和动态贝叶斯网等，由于隐马尔可夫模型的训练算法和测试算法都极为成熟，本文当中采用了隐马尔可夫作为基于时间序列的推理方法。隐马尔可夫模型的优势在于推理时使用了时间序列，充分地利用了上下文信息，但是其训练较逻辑回归复杂，无法利用过多的时间帧（否则会因联合概率较小而无法予以计算）。并且隐马尔可夫的训练数据使用了相同标签下的帧序列，即训练时所用的同一序列的帧对应的行为是相同的，而在实际过程中，同一个序列下的不同帧可能会出现不同的行为。

此外，当前的隐马尔可夫训练算法大都只针对一个离散观测量或者一个连续的随机向量的应用场景，而的观测值中同时存在着多个离散观测量和连续观测量。直观的做法是将多个离散观测量聚合成为一个单独的离散观测量，但是这种做法会使模型的参数迅速增加，例如，若在 HMM 当中选取 5 个隐状态，同时有 10 个离散观测量，每个离散观测量对应 2 个不同的取值，则观测矩阵的参数个数为 $5\times2^{10}=5120$，但是若引入朴素贝叶斯假设，即观测量之间是相互独立的，那么观测矩阵的总参数量则降为 $5\times2\times10=100$，实际中特征维度可能会更高，若不采用朴素贝叶斯假设，则由于训练样本个数较少，很难得到对模型参数合理的估计。因此，的在训练 HMM 模型时对于多维离散观测值引入了朴素贝叶斯假设。

使用 HMM 进行行为理解时，每个行为都被认为是一个序列。训练 HMM 模型时，需要利用第 j 层的公有特征集 ${}_pF_k^j$ 针对 ${}_cA^j$ 中的每个行为训练出 HMM 模型，同时需要针对每个行为训练出相应的 HMM 模型以用于第 J 次推理。在进行第 j 次推理时，利用已经观测的特征向量序列，可以计算出 ${}_cA^{j-1}$ 中的每个行为输出该序列的概率，若所得概率值超过阈值，则 a_z 会被添加至 ${}_cA^j$。

同单帧推理时一样，在第 J 次推理时，对于 ${}_ca_m^J(m \in \{1,2,\cdots,{}_cN_{j-1}\})$ 所使用的特征向量序列为公有特征向量序列与私有特征向量序列的并集。

4 实验验证

4.1 实验环境

实验环境的设置主要用于模拟人体在室内的日常行为场景，在该场景当中，需要识别出吃饭、看电视、吃水果、喝饮料、看书、喝水、使用电脑等 7 种不同的行为。视频数据的采集工作由分布在屋内的 4 套 AV800 综合采集卡以及 4 个 CCD 摄像机完成，其中集体分辨率最高可以达到 720×576，帧率可以达到 25 f/s。此外，采集卡的硬件压缩功能可以直接输出压缩格式的视频流。

实验环境当中配备了圆桌、电视、冰箱、书架、办公桌、茶几等家具，以及水果、饮料、食品等日常生活用品，前方提及的摄像机布置在房间的 4 个角落当中，分别连接至数据采集服务器当中以捕获场景当中发生的人体行为。实验环境布置如图 4 所示，其平面图如图 1 所示。

图 4 不同视角下的实验环境

Fig.4 Experiment environments under different views

4.2 实验数据集

在该数据集当中，共需要识别吃饭、看电视、吃水果、喝饮料、看书、喝水、使用电脑等 7 种不同的行为。该数据集共有 225 551 帧行为图像。

针对该行为集,使用了 2 层的推理框架对其进行推理;定义公有特征集为{人体的姿势,人体的朝向,人体的位置},其中人体的姿势有 2 个观测值,分别为站着和坐着,人体的朝向被离散成 8 个数值,人体的位置则由文献[22]中介绍的算法求出。针对不同的行为,分别定义了附着于其上的私有特征,具体内容如表 1 所示。

由于本文侧重于推理方式及效率的研究,对于如何通过对视频进行图像处理以获取所需的特征并没有进行深入探讨,本文中除了人体位置外的其他特征均是通过人工予以标注。事实上,若今后针对相应特征的视觉算法成熟后可以方便地集成到本文所提出的推理框架当中。

表 1　不同行为的私有特征

Table 1　Private features of different activities

行为	相应的私有特征	特征取值
吃饭	手中有餐具	{持有、不持有}
看电视	电视开着	{开着, 关闭}
吃水果	手中有水果	{持有、不持有}
喝饮料	手中有饮料	{持有、不持有}
看书	手接触到了书	{触碰,未触碰}
喝水	手中有水杯	{持有、不持有}
使用电脑	电脑开着	{开着,关闭}

4.3　实验结果

本文分别使用了基于逻辑回归和 HMM 的推理方式实现。在本节当中,针对这 2 种实现,均对比了未使用分层推理和使用分层推理的正确率。

4.3.1　基于逻辑回归的实验结果

使用未分层的推理实现时,将所有的特征直接用于训练得到特征-行为权重,此时针对 7 个不同的行为,均可以得到一组特征与行为的权重关系;此 7 组特征-行为权重可以用逻辑回归训练出的 1 个模型予以表示。而分层推理实现,则是先针对所有的行为使用公有特征训练得到特征-行为权重,其后针对每个行为,结合公有特征及私有特征,训练出相应的特征-行为权重,即分层推理实现时,最终得到 1 个描述公有特征与行为关系的模型以及 7 个描述私有特征和行为的模型。

使用环境上下文特征时,本文使用了三阶多项式来拟合行为概率同人体与家具相对位置的关系;由于研究的行为共涉及到 6 个不同的家具,故环境上下文总共包含 36 维的数据,故公有特征向量总维度为 38 维,全部特征向量的维度为 45 维。其中在未分层的推理当中引入环境上下文后,推理准确率得到了提高,在不使用环境上下文时准确率为 83.4%,使用环境上下文时,准确率为 85.20%。

表 2 中选取了部分特征在不同的推理层次当中对于看书的影响。从中可以看出:1)分层推理时不同层次时同一特征的权重并不相同,并且在第 2 层推理时,公有特征对行为的影响变弱,这与直观感觉相符,即公有特征在判断模糊语义时可以起到很强的作用,但是在进行确定语义的判断时则不会起到较强的作用;2)私有特征对相应行为的影响很大,这也充分证明了实验中对特征的分层是比较合理的。

表 2　特征在不同层数对看书的影响

Table 2　The influence of features on reading in different inference layers

特征	第 1 层对看书的权重	第 2 层对看书的权重
身体姿势	0.273 260	0.141 081
身体朝向	0.266 318	0.115 331
手触碰书	*N/C*	1.743 567

表 3 中所示为基于逻辑回归的不同层次推理的实验结果。从中可以得出以下结论:1)两者的推理精度接近,但是实用时分层的推理可以不用提取环境当中的所有特征,因此可以有效地节省系统效率;2)单层推理的模型总参数量为 7×45=315 个,而双层推理模型总参数量为 7×38+7×39=539,双层推理的总参数虽然更多,但是单个模型的参数量却在减少,在待识别行为集变大,且训练样本不足的情况下,可以有效地降低过拟合的可能性;3)在公有特征和私有特征划分合理的情况下,若有新的行为加入,只需要重新训练公有特征与各个行为对应的权重,以及该行为的私有特征对应的权重即可,扩展代价较小。

表 3　不同层数的推理比较

Table 3　The comparison of inference with different layers

推理层数	准确率/%	模型数量	参数数量
1	85.20	1	315
2	85.61	8	539

4.3.2　基于 HMM 实验结果

在 HMM 实现当中,对人体位置和家具位置进行离散化处理,以使其符合观测模型。未分层的 HMM 推理即在观测时使用全部的特征值进行观测序列的似然值计算,而分层的推理则首先使用公有特征对观测序列进行似然值计算,对于过阈值的行为种类再结合其私有特征进行新一轮

的似然值计算,得到最终结果。识别当中,其隐结点可能具有语义特征,但若是应用在视觉领域,则是极有可能代表一些未知的中间状态,只有最后的分类结果才是有语义的[23]。因此,在实验当中针对不同的序列长度均在不同的 HMM 当中需要注意的主要有 2 个参数:所取的观测帧数和隐结点数目。HMM 若是应用在语音隐结点数目下进行实验,实验中分别选取了 5 帧、10 帧、15 帧以及 20 帧,实验结果如图 5 所示。

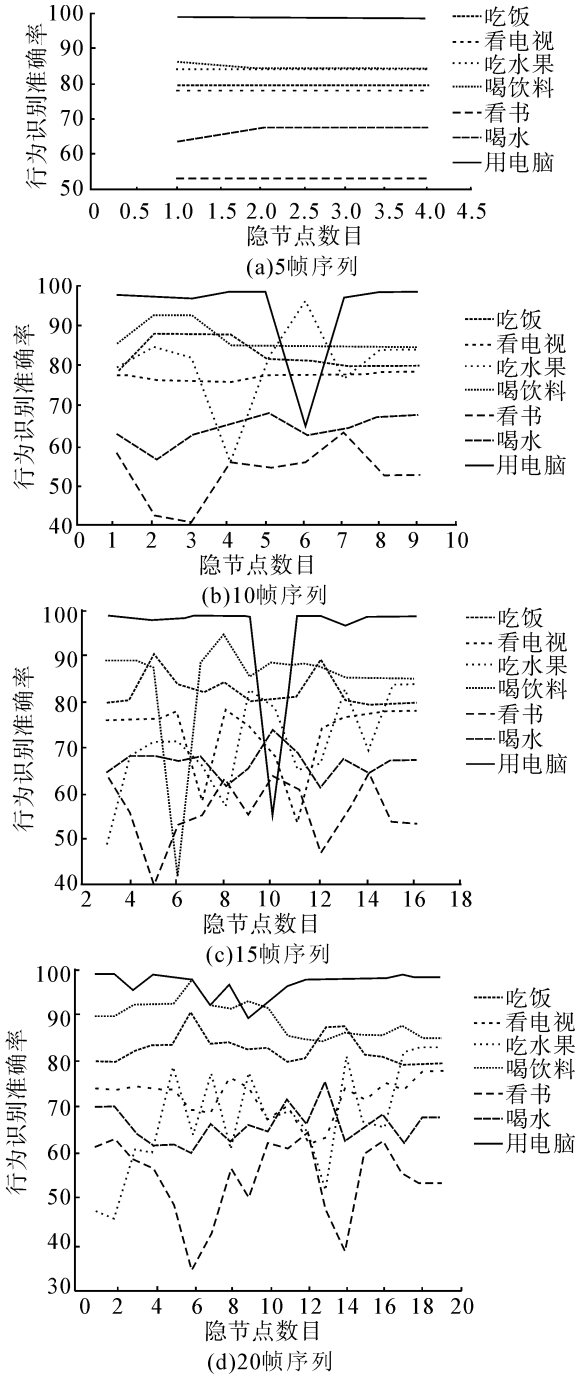


图 5 不同长度序列在不同的隐结点数目下得到的各个行为的准确率

Fig.5 Inference accuracy of different sequences under different numbers of hidden nodes

从图 5 中可以看出在序列帧数选择较少时,节点数目对识别结果几乎没有影响,而且整体识别准确率相对较低,这是因为可以使用的历史信息不够充分。而在序列帧数较多时,隐结点数目取为帧数的一半左右为宜。对于帧数为 20 的序列,2 层推理和 1 层推理的实验结果如图 6 所示。而隐结点数目为 8 的条件下(该条件下准确率相对较高)的实验结果如表 4 所示。

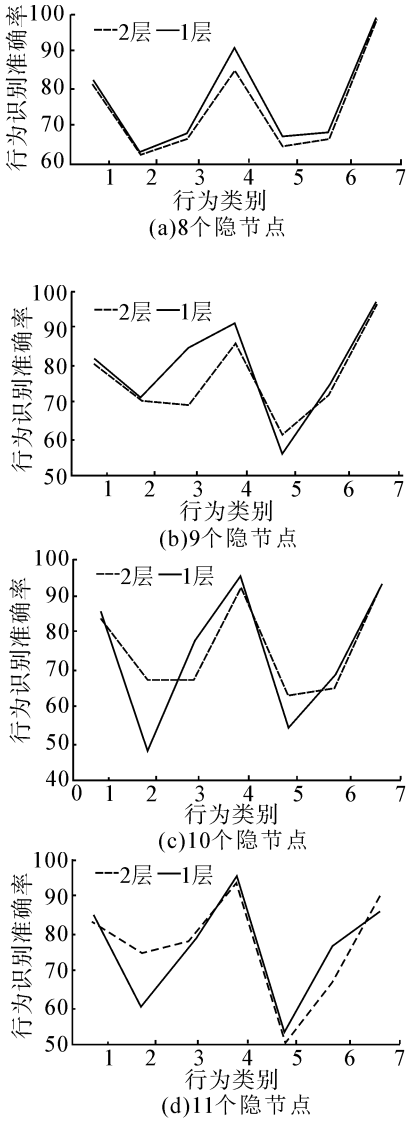


图 6 2 层 HMM 和 1 层 HMM 的实验结果对比

Fig.6 The comparison of 2-layer HMM and 1-layer HMM

可以得到类似于单帧推理的结论,并且还应该注意到对于某些行为,分层后的推理准确率会有所上升,这是因为分层推理模型当中不同层次下相同特征可以有不同的权重(不同层次下的 HMM 模型的观测概率可以不同),该行为对应的私有特征可以更好地发挥作用,而在未分层推理时,其对应的私有特征的作用可能被其他所影响。

表 4　20 帧 8 个隐节点下的各行为的准确率以及总体准确率

Table 4　Inference accuracy of different activities under the configuration of 20 frames and 8 hidden nodes

行为	2 层准确率/%	1 层准确率/%
吃饭	81.27	82.18
看电视	62.48	62.66
吃水果	66.54	67.82
喝饮料	84.70	90.99
看书	64.64	67.22
喝水	66.54	68.41
使用电脑	74.32	76.06
总体	74.32	76.06

5　结束语

本文将环境上下文作为公有特征用于行为理解,从而实现了环境上下文的可计算性,并可以对环境信息进行更加精确的描述。此外,本文将行为推理的过程分为了获取模糊语义和确定语义 2 个阶段;系统在推理过程中,根据当前观测的公有特征进行判断,筛选出模糊语义满足条件的候选行为集;如此迭代,直到依据候选行为集中的行为,观测其私有特征,并做出最为精确的判断并确定当前的语义。该框架避免了传统算法未对语义分层而提取环境中所有特征的弊病,可以有效地提升系统性能,已经在基于真实场景的数据集中得到了初步验证。

参考文献:

[1] PANTIC M, PENTLAND A, NIJHOLT A, et al. Human computing and machine understanding of human behavior: a survey[C]//Proceedings of ACM International Conference on Multimodal Interfaces. Banff, Canada, 2006: 260-266.

[2] 石为人, 周彬, 许磊. 普适计算: 人本计算[J]. 计算机应用, 2005, 25(7): 1479-1484.
SHI Weiren, ZHOU Bin, XU Lei. Pervasive computing: human-centered computing [J]. Computer Applications, 2005, 25(7): 1479-1484.

[3] 陶霖密, 杨卓宁, 王国建. 行为理解的认知方法[J]. 中国图象图形学报, 2014, 19(2): 167-174.
TAO Linmi, YANG Zhuoning, WANG Guojian. Cognitive reasoning method for behavior understanding[J]. Computer Applications, 2014, 19(2): 167-174.

[4] SHARIAT S, PAVLOVIC V. A new adaptive segmental matching measure for human activity recognition[C]//Proceedings of IEEE International Conference on Computer Vision. Sydney, 2013: 3583-3590.

[5] BOUCHARD B, GIROUX S, BOUZOUANE A. A smart home agent for plan recognition of cognitively-impaired patients[J]. Journal of Computers, 2006, 1(5): 53-62.

[6] CHEN L, NUGENT C D, MULVENNA M, et al. A logical framework for behavior reasoning and assistance in a smart home[J]. International Journal of Assistive Robotics and Mechatronics, 2008, 9(4): 20-34.

[7] THOMSON G, TERZIS S, NIXON P. Situation determination with reusable situation specifications[C]//Proceedings of IEEE International Conference on Pervasive Computing and Communications Workshops. Pisa, Italy, 2006: 620-623.

[8] ISHIMARU S, UEMA Y, KUNZE K, et al. Smarter eyewear: using commercial EOG glasses for activity recognition [C]//Proceedings of ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication. [S.l.], 2014: 239-242.

[9] HUNH T, SCHIELE B. Unsupervised discovery of structure in activity data using multiple eigenspaces[C]//Proceedings of Second International Workshop on Location-and Context-Awareness. Dublin, Ireland, 2006: 151-167.

[10] LIAO L, FOX D, KAUTZ H. Extracting places and activities from GPS traces using hierarchical conditional random fields[J]. The International Journal of Robotics Research, 2007, 26(1): 119-134.

[11] WARD J A, LUKOWICZ P, TROSTER G, et al. Activity recognition of assembly tasks using body-worn microphones and accelerometers[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(10): 1553-1567.

[12] LIU C D, CHUNG Y N, CHUNG P C. An interaction-embedded HMM framework for human behavior understanding: with nursing environments as examples[J]. IEEE Transactions on Information Technology in Biomedicine, 2010, 14(5): 1236-1246.

[13] SINGLA G, COOK D J, SCHMITTER-EDGECOMBE M. Recognizing independent and joint activities among multiple residents in smart environments[J]. Journal of Ambient Intelligence and Humanized Computing, 2010, 1(1): 57-63.

[14] WANG S, PENTNEY W, POPESCU A M, et al. Common sense based joint training of human activity recognizers [C]//Proceedings of IJCAI. Hyderabad, India, 2007:

2237-2242.

[15] SMINCHISESCU C, KANAUJIA A, METAXAS D. Conditional models for contextual human motion recognition[J]. Computer Vision and Image Understanding, 2006, 104(2): 210-220.

[16] LEE S W, MASE K. Activity and location recognition using wearable sensors[J]. IEEE Pervasive Computing, 2002, 1(3): 24-32.

[17] WANG G, JIANG J, SHI M. A context model for collaborative environment[C]//Proceedings of IEEE International Conference on Computer Supported Cooperative Work in Design. Nanjing, China, 2006: 1-6.

[18] LI M. Ontology-based Context information modeling for smart space[C]//Proceedings of IEEE International Conference on Cognitive Informatics and Cognitive Computing. Banff, Canada, 2011: 278-283.

[19] MOESLUND T B, HILTON A, KRüGER V. A survey of advances in vision-based human motion capture and analysis[J]. Computer Vision and Image Understanding, 2006, 104(2): 90-126.

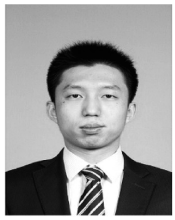
[20] AGGARWAL J K, PARK S. Human motion: modeling and recognition of actions and interactions[C]//Proceedings of IEEE International Symposium on 3D Data Processing, Visualization and Transmission. Thessaloniki, Greece, 2004. 640-647.

[21] GONZàLEZ J, VARONA J, ROCA F X, et al. aSpaces: Action spaces for recognition and synthesis of human actions[C]//Proceedings of Articulated Motion and Deformable Objects. Palma de Mallorca, Spain, 2002: 189-200.

[22] SUN L, DI H, TAO L, et al. A robust approach for person localization in multi-camera environment[C]//Proceedings of IEEE International Conference on Pattern Recognition. Istanbul, Turkey, 2010: 4036-4039.

[23] LUO Y, WU T D, HWANG J N. Object-based analysis and interpretation of human motion in sports video sequences by dynamic Bayesian networks[J]. Computer Vision and Image Understanding, 2003, 92(2): 196-216.

作者简介：



聂慧饶，男，1990 年生，硕士研究生，主要研究方向为模式识别、行为理解。

陶霖密，男，1962 年生，副教授，主要研究方向为人机交互、计算机视觉与模式识别等。承担的项目有国家重点基金情感计算，以及与 IBM、INTEL、SIEMENS 的国际合作基金等重要项目。发表论文多篇。

2015 世界机器人大会

World Robot Conference 2015(WRC 2015)

为贯彻落实习总书记在 2014 年两院院士大会上的讲话精神，积极推动创新驱动发展战略，实现我国机器人技术与产业的跨越发展，中国科学技术协会、工业和信息化部将于 2015 年 11 月在北京国家会议中心共同举办主题为“协同融合发展，引领智能社会”的 2015 世界机器人大会。

2015 世界机器人大会将由 3 项内容组成，分别是：2015 世界机器人论坛(World Forum on Robot 2015, WFR2015)、2015 世界机器人博览会(World Robot Exhibition 2015, WRE 2015)和 2015 国际青少年机器人邀请赛(World Adolescent Robot Contest 2015, WARC 2015)。

本次大会将为政府、科研机构、行业协会和和企业提供一个高端的交流平台，共同探讨、展示全球机器人的发展现状与趋势，研究机器人技术创新与产业化现状以及给我国制造业发展带来的机遇和挑战等，对促进我国机器人产业发展，推动制造业转型升级具有重要意义。

Website: http://www.cie-info.org.cn/index/tztg/201535/1425539003413_1.html