中国科学 信息科学
SCIENTIA SINICA Informationis

普适计算环境中用户意图推理的Bayes方法

易鑫, 喻纯 and 史元春

Citation: 中国科学: 信息科学 **48**, 419 (2018); doi: 10.1360/N112017-00228

View online: http://engine.scichina.com/doi/10.1360/N112017-00228

View Table of Contents:http://engine.scichina.com/publisher/scp/journal/SSI/48/4

Published by the 《中国科学》杂志社

中国科学: 信息科学 2018 年 第 48 卷 第 4 期: 419–432
SCIENTIA SINICA Informationis
智能时代的人机交互专刊 · 论文

普适计算环境中用户意图推理的 Bayes 方法

易鑫[1,3,4], 喻纯[1,2,3,4*], 史元春[1,2,3,4]

1. 清华大学计算机科学与技术系, 北京 100084
2. 清华大学全球创新学院, 北京 100084
3. 清华大学北京信息科学与技术国家研究中心, 北京 100084
4. 清华大学普适计算 (教育部) 重点实验室, 北京 100084
* 通信作者. E-mail: chunyu@tsinghua.edu.cn

收稿日期: 2017–11–06; 接受日期: 2018–01–29; 网络出版日期: 2018–04–10
国家自然科学基金 (批准号: 61521002, 61672314, 61572276)、清华大学科研基金 (批准号: 20151080408) 和网络多媒体北京市重点实验室资助项目

摘要 本文阐述了通过 Bayes 方法来预测用户交互意图的建模方法过程和推理过程. 在自然交互界面上, 用户不再是严格地通过离散明确的交互操作完成交互, 而是通过连续、非确定的多模态数据表达交互意图. 在解释用户的交互意图时, 既可以使用 "黑盒子" 的机器学习方法, 也可以利用 "白盒子" 的基于用户行为建模的方法. 后者中的用户建模, 其本质是通过计算的方法来刻画用户的行为能力, 对于理解用户意图和探索自然交互的计算原理具有重要的科学意义. 文章回顾了近年来人机交互研究中主要采用的智能算法, 向读者厘清不同方法之间的差别, 并通过我们实验室的具体研究工作展示用户建模的方法和 Bayes 推理的建模方法过程和推理过程.

关键词 Bayes 方法, 机器学习, 意图推理, 用户建模

1 交互中用户意图推理的问题及挑战

1.1 问题背景

当前, 以触摸为代表的自然交互接口已逐步取代鼠标键盘, 成为人们访问信息世界的通道. 学术界和工业界也在追求以手势、语音等多模态通道构建的自然交互界面. 自然交互的目标是让用户方便、有效地表达交互意图. 近年来, 人机交互领域持续在研究各种具有低学习成本的自然交互技术, 让计算机能够准确识别用户意图, 其中重点内容是对用户动作数据的理解和处理, 动作数据包括手指、手部、头动以及身体运动等, 是当前用户表达交互意图的主要通道.

引用格式: 易鑫, 喻纯, 史元春. 普适计算环境中用户意图推理的 Bayes 方法. 中国科学: 信息科学, 2018, 48: 419–432, doi: 10.1360/N112017-00228
Yi X, Yu C, Shi Y C. Bayesian method for intent prediction in pervasive computing environments (in Chinese). Sci Sin Inform, 2018, 48: 419–432, doi: 10.1360/N112017-00228

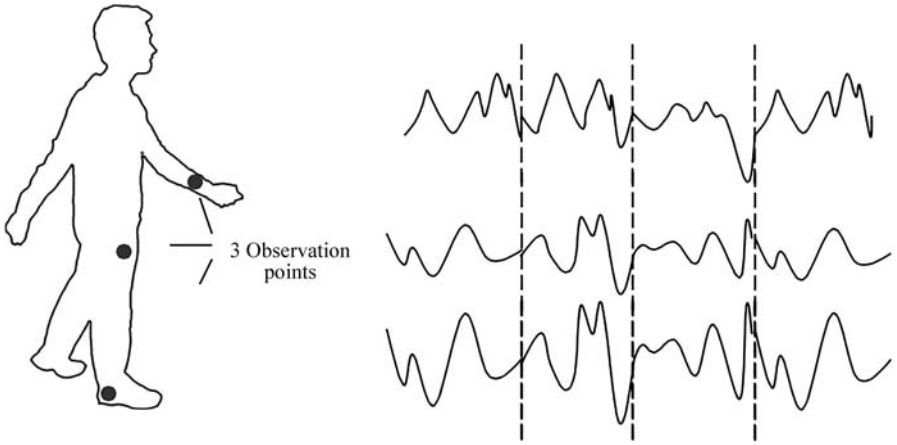

图 1　用户动作交互界面

Figure 1　Motion interaction interface

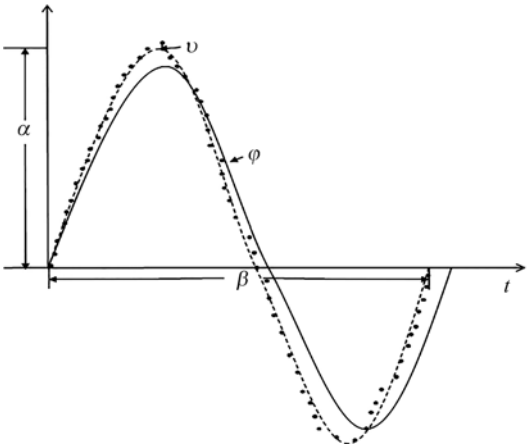

图 2　基本动作的结构

Figure 2　Structure of a basic motion

1.2　交互中的用户意图推理

动作数据处理流程可由图 1 来展示. 当用户完成一个动作时会引起界面中的观测点所采集数据的变化. 具体对于某个观测点来讲, 在交互过程中其会按照一定的时间间隔采样. 对于每次采样, 每种传感器会获得一个拥有 M 个分量的数据. 假设用户同时佩戴了 N 个独立的传感器, 即交互通道对其形成 $M \times N$ 个观测点. 同时将时间范围内所表达的交互意图记为 Y, 则这一交互过程可以形式化地表达为

$$\begin{bmatrix} S_1^1 & S_2^1 & \cdots & S_N^1 \\ S_1^2 & S_2^2 & \cdots & S_N^2 \\ \cdots & \cdots & \cdots & \cdots \\ S_1^M & S_2^M & \cdots & S_N^M \end{bmatrix} \rightarrow Y. \tag{1}$$

图 2 表示了对于一个基本动作结构的分析, 其中横轴为时间, 纵轴为数据的幅度. 根据以上分析,

可以对于基本动作构建数据产生模型. 一个基本活动可以视为由一组分段基函数的组合以一定的随机性产生, 表示为

$$f(t) = \sum \alpha_i \phi_i(\beta_i t) + \gamma_t, \tag{2}$$

其中 ϕ_i 为描述在时域上曲线变化的基函数. α_i 为幅度, 以一个随机变量表示, 并假设满足 Gauss 分布 $\alpha_i \sim N(a_i, \sigma_i^2)$. 表示动作持续时长的随机变量, 同样满足 Gauss 分布 $\beta_i \sim N(b_i, \delta_i^2)$. γ_t 表示随机误差.

对于基本动作的研究主要包含如下 3 个问题: (1) 分析以上各个组成部分所具有的性质, 从而能够对于用户的基本动作进行描述; (2) 在构建交互界面之后, 需要根据观测到的运动传感数据提取用户的交互意图, 即将信号与语义信息联系起来. 为了解决这一问题, 需要探究在更高层意图下基本动作的组成结构, 并在此基础之上进行建模和设计算法; (3) 将所提出的模型和技术应用于具体的交互问题, 包括系统预定义和用户自定义两种应用方式, 考察如何能够提高所设计界面的可用性和灵活性.

1.3 意图推理的挑战

在交互行为的意图推理中, 数据、意图和场景的变化都为推理的效果带来了挑战, 具体表现为如下 3 方面.

(1) 在完成一项动作时, 用户对自身身体运动的规划往往是确定的, 但每次做出的动作又不完全相同, 这种不确定性为交互过程中的意图推理带来了挑战. 一般情况下, 交互行为的不确定性可以归结为用户在具体完成过程中的幅度范围、时长与偏差. 本文在后续论述中, 会以动作在空间中的轨迹进行举例说明, 但是所提出的方法也可以适用于其他类型的运动数据, 如加速度计和陀螺仪等, 这些数据也可以映射为在空间中的轨迹进行分析, 具体只会在模型参数的选取方面有所不同. 另外对于数据的维度也没有限制. 以用手指在空间中画字母 'a' 为例, 幅度范围的变化体现在所画出的字母大小会有变化. 时长指的是完成整个动作在时间上的快慢, 偏差指的是在具体的完成过程中会与标准的动作形状有所差异.

(2) 不同用户针对同一交互意图可能产生不同的交互行为, 从而使得将基本动作与交互意图相互匹配更加多变和有挑战. 例如, 在利用 QWERTY 键盘进行文本输入时, 不同的用户可能采用不同的指法, 因而对于同一按键, 用户进行点击的手指是不同的, 从而基于点击手指的按键预测就需要考虑这个问题.

(3) 交互意图推理的模型需要支持不同场景下的交互行为识别. 而不同场景下用户交互行为的变化对模型的推广性带来了挑战. 纯粹基于数据的方法将由于缺乏可推广性而使应用范围大大受限. 为此, 提出的模型和技术需要从有限的数据出发, 具备一定的可解释性和可推广性, 便于灵活地适用于不同的场景.

2 现有意图推理中的智能算法

本节回顾和讨论已有工作中针对意图推理的智能算法. 如前所述, 意图推理问题的本质是一个建模问题, 即如何根据观测到的用户交互信号 S 来推断用户交互的意图 Y. 在实践中, 由于观测用户行为的方式不同 (如使用的传感器不同、观测的位置不同), 以及针对的应用场景不同, 意图推理问题的输入和输出也十分多样. 一般可以将意图推理问题分为分类和回归两类问题, 前者需要根据用户输入的信号, 从多个可能的类别中找到对应的分类 (如状态检测); 后者需要根据交互信号来计算某一具体的数值或指标, 以达到提升精度等目的. 与这些需求对应的, 意图推理的算法也有着模板匹配、决策

树、Bayes、隐变量机器学习等多种方法. 下面, 我们从技术原理的角度解释和总结近年来的意图推理工作中的智能算法, 并对其应用范围进行归纳.

2.1 模板匹配方法

模板匹配类算法是对输入进行分类的最简单方法, 其原理是针对不同的类别分别储存一个模板库, 针对每一个输入, 判断其与当前模板库中的哪个或哪些模板最相似, 即将其分类为对应的类别. 在模板匹配的过程中, 其中一个核心概念就是相似度的计算, 在实践中, 往往采用欧氏距离 [1~3].

经典的模板匹配算法假设输入信号和模板之间的序列长度是一致的, 从而将模板之间的距离简化为点对之间的距离之和 (或均值). 随着模板匹配技术被应用在越来越多的场景下, 这一 "长度相等" 的假设渐渐不满足实际使用的需求. 例如在语音交互中, 声音信号是一个时间序列, 而由于人们的语速不同, 对应于同一个音节的事件序列往往具有不同的长度. 在这些复杂情况下, 使用传统的点对距离计算法将无法求得序列之间的距离. 此时, 人们往往采用松弛的计算方法 (如 DTW 算法 [4~6]). 该方法不对点对之间的对应关系做出假设, 而是利用动态规划等方法来在所有可能的匹配中找到相似度最高的一个, 以此来作为序列间的距离. 该方法有效弥补了变长序列对距离计算带来的影响, 获得了广泛的应用.

在计算完相似度后, 由于输入的信号往往包含随机噪声, 因而若仅使用最相似的一个模板作为判断标准, 往往对噪声的鲁棒性不够. 为了改善这一点, 研究者们提出使用 k- 近邻方法 [7~12], 即计算与输入距离最接近的 k 个模板, 并取这 k 个模板中所占比最大的分类作为结果. 当 $k = 1$ 时, 就退化为最近邻方法.

模板匹配方法在分类问题的类别较少, 而且混淆性相对较小的情况下, 具有非常好的效果. 而且其计算速度很快, 因而常常被用于滑行文本输入 [2]、动作和手势识别 [1,3~6,8,10,11] 等场景中. 此外, 还有工作将其用于计算视线焦点的位置 [9]、基于屏幕触点位置判断握持姿势 [12]、身份识别 [7] 等任务. 但若模板数量过多, 或者模板之间的相似性太大, 那么这一方法的准确性和效率都会受到明显影响.

2.2 决策树和决策森林方法

决策树是一类非常简单而有效的模型, 被广泛应用在分类问题中 [13,14]. 决策树通过构造一棵树, 来对输入数据进行分层迭代的方法实现对数据的分类. 每一个叶子节点是一个决策条件, 对应着一个类别. 输入数据从根出发, 沿着对应的决策路径一路向下, 其到达的叶子节点即为其被分类的结果. 由于模型简单, 构造快速, 决策树往往具有直接的物理意义和高效的运行速度. 然而, 其缺点是在数据量不足时, 容易出现分类不准, 而在数据充足时, 又容易出现过度拟合.

针对决策树具有的限制, 人们提出了使用随机森林 [15~18] 的方法. 相比于用单一决策树, 随机森林的方法一方面提升了决策面的维度, 使得决策的准确性大大提升, 另一方面由于模型维度的大大提升, 使得该方法不仅适用于分类, 也可以用于回归问题. 目前, 决策树和随机森林的方法常被用于视线焦点位置推测 [18]、动作识别 [17]、操作手指区分 [13,15]、区分滑动操作是否正确 [16]、学习语言的规则 [14] 等任务中.

2.3 隐变量机器学习类方法

隐变量机器学习类方法往往利用包含不可见状态的随机过程或神经网络来对输入进行建模. 作为最经典的机器学习方法之一, 支持状态机 (SVM) 被广泛用于各种分类问题中 (如手势识别 [19~26]、身

份区分 [27,28]、握持姿势识别 [29]、判断是否是老人 [30]、手指是否触碰 [31]、膝盖弯曲程度 [32]), 具有计算速度快、准确性高的优点. SVM 的本质是将训练数据所在的输入空间进行划分, 以最大化不同类别之间的区分性. 针对线性不可分等复杂场景, SVM 使用不同的核函数来对输入空间进行变换, 以达到对不同数据集的分离效果.

针对基于序列数据的预测问题, 人们提出了隐 Markov 模型 (HMM). HMM 包含可被用户观测到的观测向量, 以及按照某种概率密度分布产生这些观测向量的状态序列. 利用 Markov 模型 (如 Markov 链) 对序列数据进行建模, 可以基于用于训练的观测向量数据, 得到隐含的状态序列, 以及对应的概率密度分布. 目前, HMM 被广泛用于基于事件序列的多种分类 (如手势识别 [33]、判断运动方向 [34]、是否睡着 [35]) 和回归 (提升点击位置准确性 [36]) 问题. 与之类似, Markov 决策过程和 Gauss 过程也通过随机状态链的方法来对时序交互过程建模, 用于预测未来交互行为 [37] 和推断视线焦点位置 [38].

随着神经网络的发展, 多层感知机 (MLP) [39,40] 和卷积神经网络 (CNN) [41,42] 等方法也逐渐被研究者们应用到交互意图推理中来. 神经网络通过利用多层的 “神经元” 来实现描述性极强的模型, 其在基于图像的分类和回归等问题上具有极高的准确性. 目前, 这些方法被用于基于图像和信号的身份区分 [39]、动作识别 [40,42] 和焦点位置推断 [41]. 但是, 神经网络的训练需要极其大量的训练数据, 同时其模型参数的确定也容易陷入过拟合等问题. 在以上方法之外, 隐变量机器学习类方法还包括话题模型 [43]、集成学习 [44] 等其他方法, 这里就不一一展开了.

2.4 Bayes 方法

Bayes 算法是基于统计学的一种分类算法. 它基于概率论的知识, 利用从观测数据中统计出来的特征量构建模型. 针对输入的用户数据, 该模型利用 Bayes 法则计算其归属于每一个分类的概率. 由于该方法基于概率论和统计学, 因而其模型相对比较简单, 而且模型参数一般具有直接的物理意义. 在实践中, Bayes 方法常常能实现高效的分类效率, 同时拥有不输于神经网络等方法的分类准确性. 相比于其他方法, Bayes 方法针对较小的样本, 能产生更加准确的结果, 而且能在结果之外, 同时计算出结果的置信度和显著性 [45].

在实践中, Bayes 方法常用于各种分类 (如文本输入 [2,46,47]) 和回归 (如推断用户的认知模型 [48]、提升感知深度的准确性 [49]) 问题中.

2.5 可理解性

基于以上分析, 我们可以将智能算法分为 3 类.

- 黑盒子 (如隐变量机器学习类方法): 这一类算法不需要大量的人力建模, 但强烈地依赖于训练数据. 在拥有大量的覆盖所有应用情况的数据的条件下, 其具有良好的拟合效果, 否则效果较差. 另一方面, 其训练数据需要大量的人工标注, 实际应用中有困难.
- 白盒子 (如决策树方法): 这一类算法来源于人们对客观规律的数学建模, 具有较强的推广性. 在实践中, 往往只需要基于少量的数据确定模型参数即可. 但是, 客观规律模型的建立需要耗费人力.
- 灰盒子 (如 Bayes 方法): 这一类算法结合上述两者的特点, 通过概率统计方法将人的知识引入到算法模型中, 对于无法确定的变量、关系, 通过黑盒子的方法来完成. 其同时具有对规律的可解释性和对数据的忠诚性.

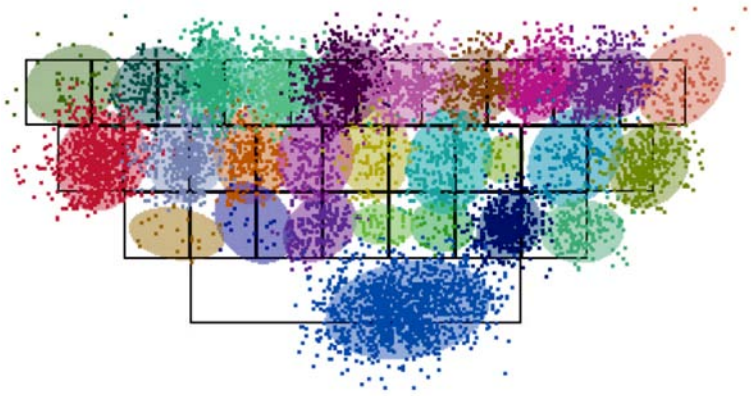

图 3 (网络版彩图) 触摸模型示例

Figure 3 (Color online) Illustration of a touch model

3 Bayes: 结合先验知识和数据的推理方法

以文本输入为例, 本文介绍 Bayes 方法如何根据用户的行为数据, 推断出用户希望输入的目标单词. 这里, 我们针对 "空中十指盲打" 的场景, 即双手悬空, 利用十个手指在物理键盘上的肌肉记忆进行空中盲打. 我们通过 Leap Motion 等传感器可以实时追踪每个手指的指尖的三维位置. 利用 Bayes 方法, 我们可以根据手指的运动规律, 以及语言本身的冗余性规律, 来对输入的目标单词进行比较准确的推测.

3.1 触摸的空间模型

触摸模型是文本输入领域一个非常重要的概念, 其定量地描述了利用某一种方法进行文本输入的过程中, 人的输入行为中的噪声的数学规律. 例如, 当在触摸屏上利用 QWERTY 键盘进行文本输入时, 由于触摸点击行为的不准确性, 人们点击一个按键时, 落点往往不是集中在同一个位置, 而是围绕着目标按键附近呈一定的分散分布, 如图 3 所示. 大量基于实际点击数据的研究表明, 对应于每一个按键的落点近似可以用一个二维 Gauss 分布 $N(\mu,\sigma^2)$ 来拟合. 其中 μ 为这些落点的中心, 以统计均值来计算; σ 为这些落点的分散程度, 以落点分布的标准差来计算. 因此, 根据 Bayes 方法的思想, 对于一个落点, 可以用触摸模型来计算出目标按键是任何字母的概率, 从而实现对用户行为的建模和解释.

在空中打字的场景中, 由于人的输入行为不是发生在触摸屏或物理键盘这样的二维平面上, 而是发生在空中, 所以用二维 Gauss 分布来对触点位置进行建模是不够的. 为此, 我们将触摸模型进行了推广, 根据中心极限法则, 针对每个按键, 我们用三维 Gauss 分布来对落点的位置进行建模. 这种建模背后仍然是 Bayes 方法的思想和计算方法, 但能很好地应对输入空间是三维空间的情况.

3.2 空中多指联动模型

在触摸模型中, 由于点击行为的噪声, 相邻按键之间的点击往往具有比较大的歧义性. 例如若一个落点刚好落在两个按键之间的边界上, 那么其目标是这两个按键的概率将可能十分接近, 从而难以做出比较好的区分. 针对这个问题, 我们利用了十指盲打的一个特点来极大地增强单次点击的信息量. 在盲打中, 按照标准指法, 相邻的两列按键 (如 'Q' 对 'W') 往往是由不同的手指点击. 那么, 如果能利用手指的信息, 将会对落点位置的解释带来极大的帮助. 例如, 若落在 'Q' 和 'W' 之间的落点是由小

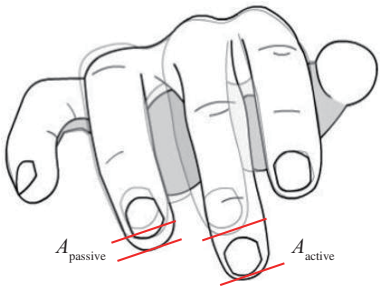

图 4 (网络版彩图) 当点下中指时, 由于手指联动, 无名指也会向下运动

Figure 4 (Color online) When tapping the middle finger, the index finger moves along it due to the correlation movement between fingers

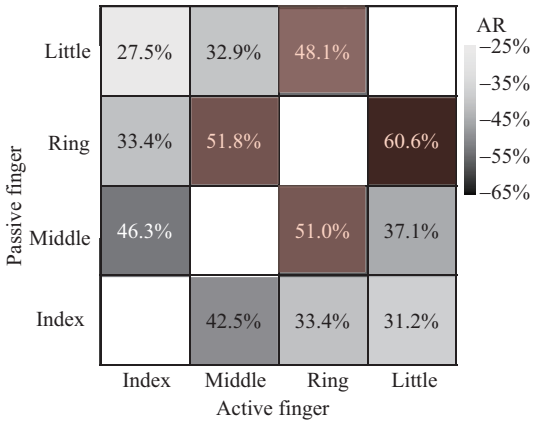

图 5 (网络版彩图) 协同运动手指的幅度

Figure 5 (Color online) The amplitude ratio (AR) of correlated finger movement

拇指触发的, 那么其应该有更大的概率是 'Q', 而同一位置的落点如果由无名指触发, 那么其应该有更大的概率是 'W'.

基于上面的观察, 我们希望能在触摸模型中不仅考虑落点的位置, 还同时考虑触发落点的手指. 然而, 在空中打字时由于缺少物理的平面和对应的触觉反馈, 人们往往会出现手指的联动现象. 即在按下一个手指时 (如中指), 由于手的生理结构, 也会导致其他手指 (如无名指) 同时向下按下. 联动现象导致我们经常观察到的是多个手指的同时按下, 从而使我们无法准确地区分出当前是哪个手指在触发点击. 图 4 展示了无名指与中指之间的联动现象.

为了应对这个问题, 我们仍然采用 Bayes 方法的思路, 针对观察到的一次多指点击, 我们分别计算它是每个触发手指的概率. 为了计算这个概率, 我们首先通过实验观察到一个结果: 当主动手指点击下时, 不同随动手指的点击幅度是不同的, 而且具有一个固定的规律. 例如, 当无名指点下 1 cm 时, 中指平均会同时下移 0.51 cm, 而食指平均会下移 0.33 cm. 据此, 我们利用一维 Gauss 模型对随动手指的移动幅度进行了建模, 如图 5 所示, 从而可以用该模型计算单次点击是各个手指的概率.

3.3 增强的 Bayes 模型

综合以上方面, 我们利用了三维的空间触摸模型和空中多指联动模型, 从 Bayes 方法的角度对用户的打字行为进行了建模和解释. 至此, 我们已经可以基于用户的打字行为来推测其目标按键了. 不过, 这样的计算仅仅利用了对用户行为的理解, 而完全没有用到语言本身的信息. 例如, 在目标语言中不可能发生的字母组合 (非单词), 以及不同单词本身出现的概率不同 (词频) 等. 基于这些信息, 我们可以进一步增强我们的 Bayes 模型, 实现更加准确的输入理解效果.

具体而言, 我们的输入是用户产生的一系列落点 I, 以及产生这些落点时, 每个手指的点击幅度 D. 基于 Bayes 思想, 我们需要计算的是 $P(W|I,D)$, 即在给定 I 和 D 的情况下, 输入的目标单词是 W 的概率, 其中 W 是任意一个可能的单词.

基于 Bayes 公式, 该条件概率正比于 $P(W)P(I|c)P(D|c)$. 其中 $P(W)$ 是语言模型, 衡量了 W 在当前环境中出现的概率. 在实践中, 一般使用其词频来计算. $P(I|c)$ 衡量了用户的目标按键和落点实际

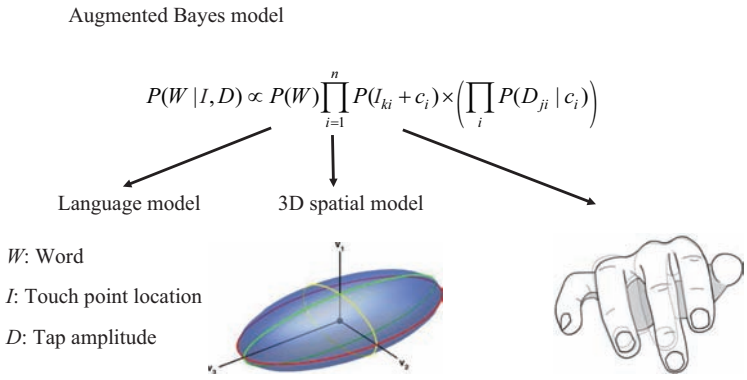

图 6 (网络版彩图) 增强的 Bayes 模型

Figure 6 (Color online) Augmented Bayes model

位置之间的概率关系, 可以用三维空间触摸模型来计算. $P(D|c)$ 衡量了用户的目标按键和使用的主动手指之间的概率关系, 可以用空中多指联动模型来计算 (见图 6).

通过结合语言模型、触摸模型和空中多指联动模型, 我们的增强 Bayes 预测模型同时考虑了落点的位置信息、手指的指法信息和语言本身的信息, 从而实现了对用户输入意图的较为准确的解释和预测, 支持了较好的空中双手盲打文本输入体验.

3.4 其他技术举例

(1) 虚拟现实 (VR) 头盔上头部运动控制规律, 及基于 Bayes 方法的单词预测. 随着 VR 头盔的不断普及, 利用头部运动进行交互也因不依赖于双手, 而在研究中成了一大热门的交互方式. 作者研究了通过头动在虚拟现实中进行文本输入的运动控制规律. 文本输入过程中, 用户通过头部运动控制光标在虚拟键盘上进行频繁的按键选择或滑动输入操作. 作者通过研究发现, 用户头部在竖直方向和水平方向上具有不同的运动控制能力. 在竖直方向上运动范围较窄, 但是目标点的控制精度较高; 水平方向上运动范围较宽, 但是控制精度较低. 针对这一发现, 作者对经典的滑动输入键盘算法进行了改进, 得到了新的算法. 在文本输入过程中, 该算法利用二维 Gauss 分布对每个按键的落点分布规律进行建模, 通过考虑竖直和水平方向上运动能力的差异性, 采用的 Gauss 分布在水平和竖直方向的参数也有差别. 并通过 Bayes 方法, 计算滑动轨迹与各单词模板轨迹之间的匹配性, 最终与单词的词频相结合, 进行输入文本预测. 相比于原有算法, 新算法有了 10% 的精度提升, 同时, 用户的输入速度也提高了 20%.

(2) 用户手腕旋转实现单手交互的运动控制能力建模. 随着智能手表等可穿戴设备的发展, 单手交互的需求越来越受到研究者的关注 (例如另一只手正提着东西). 针对这个问题, 作者研究了利用旋转手腕来在智能手表上进行指点的交互技术. 该技术通过陀螺仪感知用户手腕的转动方向, 将其映射为手表屏幕上的指针位置, 并通过确认操作完成指点任务. 作者着重研究了手腕各方向旋转的舒适范围和屏幕可视范围, 并针对精细目标的选择进行了实验, 研究了手腕旋转行为在不同方向的控制精度. 基于该研究结果, 作者提出了一种利用 Bayes 方法进行精细目标选择的技术, 该技术利用二维 Gauss 分布对用户选择目标的误差进行建模, 并结合手腕旋转的运动能力确定分布参数, 最终利用 Bayes 推断来识别用户想要点击的目标. 该研究证明, 手腕转动可以作为一种实际的输入方式解决单手交互问题.

(3) 盲输入行为模式和 Bayes 优化技术. 近年来, VR 头盔显示器逐渐流行, VR 头盔提供的沉浸感是现有的显示技术所不具备的, 因而被看成是下一代可视媒体的重要载体. 然而与之对比的是, 在 VR 头盔上还没有高效的文本输入方法, 通过头动或者通过鼠标来控制虚拟键盘上的光标, 从而实现文本输入的方法效率都不高. 因此, 作者针对该问题, 创新性地提出了盲式输入技术. 其核心思想是利用用户在手机上长期进行文本输入所获得的对键盘布局的肌肉记忆来完成输入. 作者首先对盲式状态下进行文本输入过程中, 手指的运动控制能力进行了研究, 并首次发现了连续点击的两个落点之间的位置相关性. 基于该发现, 作者创新性地提出了基于连续触点之间位置偏差的文本预测算法. 该方法可以有效地解决盲式输入过程中由于没有显示的键盘, 而导致用户想象中的键盘位置不确定的问题. 基于连续落点之间的相关性, 作者在利用 Bayes 方法进行输入文本预测的过程中, 将现有工作中 “每个落点独立用 Gauss 分布” 建模进行了扩展, 转而针对 “每个落点相对于” 上一个落点的位置偏差进行建模, 使得预测准确率得到了提升. 在实际使用过程中, 用户通过短暂的学习就可以适应盲式输入模式, 输入速度可以达到 20 单词每分钟, 接近于日常在手机上的输入速度.

(4) 交互任务的量化评价, 文本输入方法的评测方法. 对于文本输入单词评测集的研究本质是对交互任务的研究, 是解释用户交互动作的场景和前提. 文本输入的评测集是评价文本输入算法效果的重要的工具. 一般情况下, 研究者在实验中让用户对照文本评测集中的语句进行输入, 同时统计输入的速度和错误率来对文本输入技术进行评测. 目前, 文本评测集是基于经验选择出来的. 在我们的研究过程中, 我们发现部分单词相对于其他单词更加 “容易” 输入. 这里, “容易” 是因为这些单词不易与其他单词混淆, 因此, 即便用户输入过程中存在一定程度的错误, 也可以比较简单地通过输入法的智能算法进行纠正. 为此, 我们认为文本评测集中的句子本身是否 “容易” 输入应该是在选择评测集时需要考虑的一项重要指标, 否则会导致性能测试结果虚高或虚低. 针对该问题, 我们专门定义了单词清晰度的概念, 用以量化一个单词与其他单词发生混淆的难易程度. 我们系统地研究了单词清晰度的计算方法, 并且对目前常用评测集中的短语进行了优化, 生成了一系列具有合理单词清晰度分布的语料库. 作者提出的单词清晰度概念直接由 Bayes 方法推导得到, 其物理意义为 “用户意图输入该单词时, 其实际产生的行为匹配于其他单词的概率”. 因而, 基于单词清晰度概念优化得到的评测集, 可使利用 Bayes 方法进行输入文本预测的技术评测更加公平和有效.

4 Bayes 方法的适用范围和局限, 以及未来工作

对交互的意图推理包含两个方面: 一是交互情境和状态的推理, 用户的交互行为一定是发生在某一情境和状态下的, 不同情境下的同一交互行为可能表达了不同的交互意图. 因而, 对交互情境的推理是实现准确交互意图推理的前提; 二是用户交互意图的推理, 在一定情境下, 用户的交互意图会通过其状态的变化或者某些特定的行为体现出来. 因而, 如何根据对用户的观测来准确推测交互意图, 是一个十分重要的问题.

4.1 适用范围

相比于更加直接的模板匹配类方法, 以及更加复杂的机器学习类方法, 作为 “灰盒子” 的 Bayes 方法具有两者融合的优势. 一方面, 其对训练数据的规模要求较小, 可以在样本量不多时就产生较好的结果. 而与之对比的, 机器学习类方法需要大量的训练数据以实现较好的效果, 模板匹配类方法也由于建模过于简单而容易受样本噪声等因素的影响. 与之对比, Bayes 方法在给出单一的分类或回归结果

之外, 还能计算出结果的置信度, 因而对于移动、可穿戴这些交互行为模糊、数据包含噪声的场景, 其对交互意图的推断有着更加丰富的适用性.

另一方面, 得益于 Bayes 方法中的人工建模, 我们可以在一定程度上把握意图推理问题的核心规律. 因而, 其模型参数往往具有直接的概率或物理意义, 参数的取值也因此更容易从数据中通过统计等简单方法获得. 在个性化和迁移到不同场景等需求中, 用户的数据往往具有某一类似的特性, 但随着时间的变化或在不同的场景中, 在具体数值上有一定的差别. 在这种情况下, 隐变量类机器学习类方法在选定模型结构后, 其训练结果完全由训练数据决定, 因而针对新交互情境下的交互行为, 往往难以通过参数的微调实现模型的迁移, 而需要重新训练, 从而带来可观的人力和计算开销. 而与之对比, Bayes 方法则可以在保持核心建模不变的前提下, 通过参数的动态调整得到适应于新用户或新场景的模型.

4.2 Bayes 方法的局限

Bayes 方法虽然具有如上的优势, 但其本身仍然具有一定的适用范围和局限性. 首先, 如上所述, Bayes 方法的优势在特定的交互环境中才有明显的体现, 例如移动、可穿戴环境等. 在这些环境下, 用户的交互行为数据往往具有模糊性, 且常常包含噪声. 例如, 在空中利用手势交互时, 不同手势之间常常具有一定的相似性, 因此难以根据数据进行确定的分类. 在触屏文本输入中, 人们点击单个按键时常常产生偏差, 而 Bayes 方法则可以将每次点击的不确定度量化, 进而利用整个输入过程的信息推测目标单词. 但是, 对于判断用户的输入状态等任务而言, Bayes 方法的效果就会受到限制. 例如, 若要判断用户是否睡着、手机的持握姿势、用户的身份区分等, 系统可能针对不同的情境有着十分不同的行为, 此时就无法通过 Bayes 方法来进行概率化的判断, 而需要使用机器学习等方法来进行更准确的确定性判断.

另一方面, Bayes 方法由于需要人工建模, 因而其受限于人对问题本质的把握能力, 模型的复杂度也是有限的. 与此对应, Bayes 方法更适用于推理相对浅层的交互意图, 如眼睛的焦点位置、文本输入的目标单词等, 而相对深层的交互意图则难以被人工建模来描述, 如推断认知模型、预测未来交互行为等. 此外, 基于图像等相对复杂的输入信号时, 人们往往也需要利用更高维度的模型来进行意图推理, 此时, 机器学习类方法就可以发挥对应的优势.

4.3 未来工作

针对以上限制, 我们的未来工作可以从如下两个方面展开.

(1) 在更多交互场景下, 挖掘和验证 Bayes 方法的适用性. 当前, 随着移动、可穿戴、空中交互等新兴交互情景的兴起, 人们在这些情景下的交互行为也越来越自然, 随之而来的就是交互信息的模糊化. 与之对应的 Bayes 方法可以很好地在这样的条件下对交互意图进行推理. 因而, 我们可以在更多场景中对 Bayes 方法的适用性进行验证. 例如, 在虚拟现实环境中, 用户利用身体姿态和手势实现对虚拟内容的交互, 而利用运动传感器、摄像头信号所实现的动作追踪和识别就成为其中关键的技术问题. 此外, 在针对盲人的触摸屏交互中, 如何利用语音提供的交互反馈来尽可能地增加盲人在虚拟键盘上的文本输入效率和舒适感, 也是可以利用 Bayes 方法进一步探索的问题.

(2) 针对用户深层次意图的挖掘, 可以通过 Bayes 方法和机器学习等方法的融合, 提升传统 Bayes 方法的建模能力, 达到兼具两者优势的效果. 例如, 在手机屏幕上基于电容屏信号的用户触摸意图推测

中, 可以先由机器学习方法实现对显著特征的提取和初步处理, 实现输入信号的综合和降维, 接着利用 Bayes 方法, 基于处理后的输入数据进行意图推断, 以提出可解释和可移植的模型.

5 总结

本文讨论了 Bayes 方法在用户交互意图推理中的应用. 通过对已有工作中的智能方法的调研和分析, 我们比较了各种智能方法的特性和适用范围, 分析了 Bayes 方法相对于模板匹配和机器学习类方法的优势. 接着, 我们以空中双手盲打文本输入情境为例, 介绍了 Bayes 方法的实际使用方法和效果, 并以一系列其他技术作为案例支持. 最后, 我们讨论了 Bayes 方法的适用范围、局限性和未来工作. 在普适计算环境下, 新兴的交互模态和界面层出不穷, 对应于每一种具体的情景和交互任务, 都有不同的输入、输出和建模需求. 虽然 Bayes 方法在大部分输入信号具有模糊性、交互意图不深的情况下具有良好的效果, 但实际使用中, 研究者们还是需要针对具体的情景和需求, 灵活地使用不同的方法, 甚至对不同的方法进行融合, 以达到最好的交互效果.

参考文献

1 Laput G, Zhang Y, Harrison C. Synthetic sensors: towards general-purpose sensing. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, 2017. 3986–3999

2 Yu C, Gu Y Z, Yang Z C, et al. Tap, dwell or gesture?: exploring head-based text entry techniques for HMDs. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, 2017. 4479–4488

3 Ilisescu C, Kanaci H A, Romagnoli M, et al. Responsive action-based video synthesis. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, 2017. 6569–6580

4 Wu C-J, Houben S, Marquardt N. EagleSense: tracking people and devices in interactive spaces using real-time top-view depth-sensing. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, 2017. 3929–3942

5 Vatavu R-D. Improving gesture recognition accuracy on touch screens for users with low vision. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, 2017. 4667–4679

6 Taranta II E M, Samiei A, Maghoumi M, et al. Jackknife: a reliable recognizer with few samples and many modalities. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, 2017. 5850–5861

7 Schneegass S, Oualil Y, Bulling A. SkullConduct: biometric user identification on eyewear computers using bone conduction through the skull. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, 2016. 1379–1384

8 Liu M Y, Nancel M, Vogel D. Gunslinger: subtle arms-down mid-air interaction. In: Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology, Charlotte, 2015. 63–71

9 Sugano Y, Bulling A. Self-calibrating head-mounted eye trackers using egocentric visual saliency. In: Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology, Charlotte, 2015. 363–372

10 Huang D, Zhang X Y, Saponas T S, et al. Leveraging dual-observable input for fine-grained thumb interaction using forearm EMG. In: Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology, Charlotte, 2015. 523–528

11 Sun K, Wang Y T, Yu C, et al. Float: one-handed and touch-free target selection on smartwatches. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, 2017. 692–704

12 González R M, Appert C, Bailly G, et al. TouchTokens: guiding touch patterns with passive tokens. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, 2016. 4189–4202

13 Gil H, Lee D Y, Im S, et al. TriTap: identifying finger touches on smartwatches. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, 2017. 3879–3890

14 Hanafi M F, Abouzied A, Chiticariu L, et al. SEER: auto-generating information extraction rules from user-specified examples. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, 2017.

6672–6682
15 Sridhar S, Markussen A, Oulasvirta A, et al. WatchSense: on- and above-skin input sensing through a wearable depth sensor. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, 2017. 3891–3902
16 Noor M F M, Rogers S, Williamson J. Detecting swipe errors on touchscreens using grip modulation. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, 2016. 1909–1920
17 Chan L W, Chen Y-L, Hsieh C-H, et al. CyclopsRing: enabling whole-hand and context-aware interactions through a fisheye ring. In: Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology, Charlotte, 2015. 549–556
18 Huang M X, Kwok T C K, Ngai G, et al. Building a personalized, auto-calibrating eye tracker from user interactions. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, 2016. 5169–5179
19 Krupka E, Karmon K, Bloom N, et al. Toward realistic hands gesture interface: keeping it simple for developers and machines. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, 2017. 1887–1898
20 Laput G, Xiao R, Harrison C. ViBand: high-fidelity bio-acoustic sensing using commodity smartwatch accelerometers. In: Proceedings of the 29th Annual Symposium on User Interface Software and Technology, Tokyo, 2016. 321–333
21 Chen X, Li Y. Bootstrapping user-defined body tapping recognition with offline-learned probabilistic representation. In: Proceedings of the 29th Annual Symposium on User Interface Software and Technology, Tokyo, 2016. 359–364
22 Zhou J H, Zhang Y, Laput G, et al. AuraSense: enabling expressive around-smartwatch interactions with electric field sensing. In: Proceedings of the 29th Annual Symposium on User Interface Software and Technology, Tokyo, 2016. 81–86
23 Zhang Y, Xiao R, Harrison C. Advancing hand gesture recognition with high resolution electrical impedance tomography. In: Proceedings of the 29th Annual Symposium on User Interface Software and Technology, Tokyo, 2016. 843–850
24 Zhang Y, Harrison C. Tomo: wearable, low-cost electrical impedance tomography for hand gesture recognition. In: Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology, Charlotte, 2015. 167–173
25 Lin J-W, Wang C, Huang Y Y, et al. BackHand: sensing hand gestures via back of the hand. In: Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology, Charlotte, 2015. 557–564
26 Li H C, Brockmeyer E, Carter E J, et al. PaperID: a technique for drawing functional battery-free wireless interfaces on paper. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, 2016. 5885–5896
27 Holz C, Knaust M. Biometric touch sensing: seamlessly augmenting each touch with continuous authentication. In: Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology, Charlotte, 2015. 303–312
28 Li H C, Zhang P J, Moubayed S A, et al. ID-Match: a hybrid computer vision and RFID system for recognizing individuals in groups. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, 2016. 4933–4944
29 Yoon D, Hinckley K, Benko H, et al. Sensing tablet grasp + micro-mobility for active reading. In: Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology, Charlotte, 2015. 477–487
30 Hagiya T, Horiuchi T, Yazaki T. Typing tutor: individualized tutoring in text entry for older adults based on input stumble detection. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, 2016. 733–744
31 Zhang Y, Zhou J H, Laput G, et al. SkinTrack: using the body as an electrical waveguide for continuous finger tracking on the skin. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, 2016. 1491–1503
32 Leong J, Parzer P, Perteneder F, et al. proCover: sensory augmentation of prosthetic limbs using smart textile covers. In: Proceedings of the 29th Annual Symposium on User Interface Software and Technology, Tokyo, 2016. 335–346
33 Alaoui S F, Françoise J, Schiphorst T, et al. Seeing, sensing and recognizing laban movement qualities. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, 2017. 4009–4020
34 Qian K, Wu C S, Zhou Z M, et al. Inferring motion direction using commodity wi-fi for interactive exergames. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, 2017. 1961–1972

35 Fridman L, Toyoda H, Seaman S, et al. What can be predicted from six seconds of driver glances? In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, 2017. 2805–2813
36 Buschek D, Alt F. ProbUI: generalising touch target representations to enable declarative gesture definition for probabilistic GUIs. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, 2017. 4640–4653
37 Banovic N, Buzali T, Chevalier F, et al. Modeling and understanding human routine behavior. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, 2016. 248–260
38 Huang M X, Li J J, Ngai G, et al. ScreenGlint: practical, in-situ gaze estimation on smartphones. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, 2017. 2546–2557
39 Evans A C, Davis K, Fogarty J, et al. Group touch: distinguishing tabletop users in group settings via statistical modeling of touch pairs. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, 2017. 35–47
40 McIntosh J, Marzo A, Fraser M, et al. EchoFlex: hand gesture recognition using ultrasound imaging. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, 2017. 1923–1934
41 Sugano Y, Zhang X C, Bulling A. AggreGaze: collective estimation of audience attention on public displays. In: Proceedings of the 29th Annual Symposium on User Interface Software and Technology, Tokyo, 2016. 821–831
42 Wang S W, Song J, Lien J, et al. Interacting with soli: exploring fine-grained dynamic gesture recognition in the radio-frequency spectrum. In: Proceedings of the 29th Annual Symposium on User Interface Software and Technology, Tokyo, 2016. 851–860
43 Vaccaro K, Shivakumar S, Ding Z Q, et al. The elements of fashion style. In: Proceedings of the 29th Annual Symposium on User Interface Software and Technology, Tokyo, 2016. 777–785
44 Liu C, Clark G D, Lindqvist J. Where usability and security go hand-in-hand: robust gesture-based authentication for mobile systems. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, 2017. 374–386
45 Kay M, Nelson G L, Hekler E B. Researcher-centered design of statistics: why Bayesian statistics better fit the culture and incentives of HCI. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, 2016. 4521–4532
46 Yi X, Yu C, Zhang M R, et al. ATK: enabling ten-finger freehand typing in air based on 3D hand tracking data. In: Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology, Charlotte, 2015. 539–548
47 Yu C, Sun K, Zhong M Y, et al. One-dimensional handwriting: inputting letters and words on smart glasses. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, 2016. 71–82
48 Kangasrääsiö A, Athukorala K, Howes A, et al. Inferring cognitive models from data using approximate Bayesian computation. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, 2017. 1295–1306
49 Finnegan D J, O'Neill E, Proulx M J. Compensating for distance compression in audiovisual virtual environments using incongruence. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, 2016. 200–212

Bayesian method for intent prediction in pervasive computing environments

Xin YI[1,3,4], Chun YU[1,2,3,4*] & Yuanchun SHI[1,2,3,4]

1. *Department of Computer Science and Technology, Tsinghua University, Beijing* 100084*, China*;
2. *Global Innovation eXchange Institute, Tsinghua University, Beijing* 100084*, China*;
3. *Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing* 100084*, China*;
4. *Key Laboratory of Pervasive Computing, Ministry of Education, Tsinghua University, Beijing* 100084*, China*
* Corresponding author. E-mail: chunyu@tsinghua.edu.cn

Abstract This paper describes the principle and examples of predicting users' intent using the Bayesian method. In natural interfaces, users use continuous, undetermined multi-modal data to express their interaction intention rather than using discrete, determined actions. In order to interpret their interaction intend, we can either use "black-box" machine-learning methods or use "white-box" user-behavior-modelling methods. The essence of the latter is to computationally model the users' interaction ability, which is crucial to understanding the users and exploring the computation principle of natural interaction. This paper summarizes the popular intelligent algorithms used in recent HCI researches, discusses the difference between these methods, and illustrates the methods of user modelling and the Bayesian method using some works from our laboratory.

Keywords Bayes method, machine learning, intent prediction, user modelling

Xin YI is a senior Ph.D. candidate at Tsinghua University. He received his bachelor's degree from Tsinghua University in 2013. His research interests are focused on user behavior modelling and algorithms including basic interaction tasks (e.g., touch and pointing) and text entry techniques in various modalities (e.g., smartphones, smartwatches, and VR/AR).

Chun YU is an associate researcher at Tsinghua University. He received his Ph.D. degree from Tsinghua University in 2012. He is keen to research computational models and algorithms that can facilitate the development of a high-efficiency multi-modal user interface such as mobile phones, large displays, and VR/AR headsets.

Yuanchun SHI is a Changjiang distinguished professor of the Department of Computer Science, the director of HCI & Media Integration Institute. She is the dean of Global Innovation eXchange (GIX) Institute of Tsinghua University. Her research interests include human computer interaction, pervasive computing, distributed multimedia processing and e-learning.