

自然文本输入中的动作建模

喻纯 易鑫 史元春
清华大学

关键词：自然文本输入 运动模型 意图推理

引言

人机界面是人与计算机之间传递、交换信息的桥梁。几十年来，人机界面的发展越来越强调交互的自然性，即用户的交互行为与其生理和认知的习惯相吻合。

人机交互的方式经历了命令行、图形用户界面、触摸交互和三维空中交互的演变。其结果是，交互的自然性逐渐提高，但由于交互接口尺寸的限制和触觉等反馈信道受限，导致交互信号的精度和交互效率逐渐降低。这种自然性和高效性之间的制约关系，成为了人机交互研究中的难题，如何在两者之间取得兼顾和平衡，是具有重要理论价值和实践意义的研究问题。

文本输入是指人通过人机界面向计算机输入文本信息的过程，是最基本的人机交互任务之一。在日常生活中，历史最为悠久和广为接受的文本输入方式是利用键盘。传统的键盘包括物理键盘和较大尺寸的软键盘，用户经过一段时间的练习后，基本上可以实现无错的文本输入。

然而，在可穿戴设备、虚拟/增强现实等新一代自然交互场景中，用户往往面临着交互接口尺寸极小(智能手表)、缺乏触觉反馈(空中交互)等挑战，这使得用户难以保证输入的准确性，输入信号的信噪比较低。此时，传统的纠错能力弱的文本输入技术的输入正确率显著下降，最终导致其无法完成文本输入任务，或者在输入过程中导致用户紧张疲劳，输入效率低，输入体验差。

文本输入中的意图推理

为了解决自然文本输入的准确率低、输入信号的信噪比低的问题，研究者和业界开始逐渐探索从有噪声的文本输入信号中对输入意图进行推理的方法。其中，最常用的是古德曼 (Goodman) 等人提出的经典统计解码方法^[1]，其本质是贝叶斯方法。该算法体现为一个概率模型，它预先维护一个包含所有合法单词的词库，在输入过程中，算法将用户的输入视为一个包含噪声的随机信号，并在给定该信号的前提下，计算词库中每个单词的似然度，并且将那些似然度最高的单词作为输入识别结果。文本输入的意图推理问题可被形式化地定义为：设 I 是输入信号的序列， W 是词库中任意一个候选单词，算法需要寻找的是使 $P(W|I)$ 最大的 W 。根据贝叶斯公式，我们有

$$P(W|I) \propto P(I|W) \times P(W) \quad (\text{公式1})$$

在研究中， $P(W)$ 称为语言模型，量化了不同单词出现的概率。在实际应用中，该概率值可以利用多种不同的模型进行估计，如 N 元词频或基于神经网络的语言模型等； $P(I|W)$ 称为点击模型，量化了输入信号中的噪声。该贝叶斯方法同时考虑了用户输入行为的先验知识和语言本身的信息。目前，该算法已经被证明在很多智能键盘技术中都有效^[2,3]。在实践中，很多键盘算法假设用户在输入时不会产生多输、漏输、交换等错误^[2,3]，因而输入落点和目标字符之间可以建立明确的一一映射关系。进一步的，人们常常假设落点之间是相互独立的。在这种

情况下,我们可以将计算进行简化,得到:

$$P(I|W)=\prod_{i=1}^n P(I_i|W_i) \quad (\text{公式2})$$

其中 n 是输入落点的数量。此时, $P(I|W)$ 的计算就分解为对每次点击独立计算概率,并将结果相乘。

为了更好地理解用户在键盘上的输入行为,以及为点击模型提供经验参数,研究者们针对多种输入接口上的文本输入行为进行了建模,其结果一般称为打字模式。值得一提的是,绝大部分工作针对的都是触屏上的 QWERTY 软键盘这一主要的接口形式。由于人们在触屏上的点击行为具有不准确性,所以,打字模式本质上定量描述了触摸落点分布的统计性规律^[4,5]。在研究中,人们发现在点击每个按键时,落点位置在空间上符合二维高斯分布^[1,4],从而可以使用该分布的参数来计算公式2中的 $P(I_i|W_i)$ 。

然而,已有工作未能面向新型自然文本输入接口对人的输入行为进行建模,从而在输入噪声较大、反馈信道受限等条件下,点击模型与用户输入行为的吻合程度不高,模型精度不高;面向三维空中输入等接口形态,在落点分布规律之外,点击模型也未能对人的运动规律有更全面的理解,从而缺乏足够的先验知识。针对以上限制,本文将重点讨论自然文本输入中的动作建模问题。

触屏软键盘上的点击模型精度优化

触屏软键盘是日常生活中最常见的输入界面之一,随着可穿戴设备的普及,智能手表上的文本输入需求也日益强烈。然而,智能手表触摸屏幕的输入界面尺寸非常有限,对应的软键盘按键尺寸极小,几乎不可能精确瞄准和用手指点击。为了保证自然舒适的文本输入体验,用户往往需要在不仔细瞄准的情况下进行快速的点击操作,而这会导致输入落点的精度和信号的信噪比进一步降低。

针对这一文本输入接口上的“胖手指”难题,首先需要建立接口尺寸与输入速度、准确率和落点

分布的作用关系,从而提升软键盘输入的贝叶斯方法中点击模型与输入行为的吻合程度。为了覆盖主要的智能手表尺寸,我们研究了宽度从2.0cm到4.0cm的不同尺寸全键盘。通过用户实验,我们发现用户在使用不同尺寸的键盘输入时,倾向于保持相对稳定的输入速度,而落点位置的准确率则随着键盘尺寸的减小而显著降低。此外,我们还发现用户手指的触摸姿势会随着键盘的尺寸而变化,当键盘较大时,用户会倾斜手指进行触摸,而键盘较小时,用户会将手指竖直触摸,这些都导致了落点在系统偏差和分布集中程度上随着键盘尺寸的变化(见图1)。

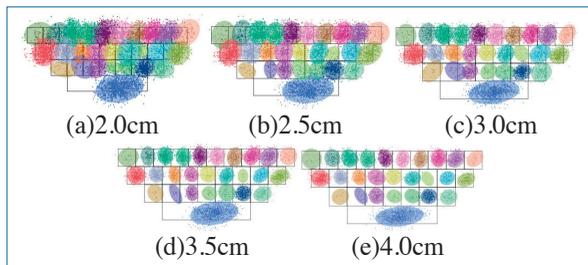


图1 不同尺寸键盘上采集的落点分布,以及95%置信椭圆

基于该结果,我们优化了文本输入的基本贝叶斯方法中的点击模型。由于超小尺寸键盘上用户的点击行为与在智能手机上不一样,因而落点的分布规律也有差别,从而无法将现有的智能手机软键盘的点击模型参数直接应用在该场景中。实验证明,在结合了优化精度的点击模型和单元语言模型的贝叶斯算法支持下,用户可以在宽度仅为3.5cm的智能手表全键盘上达到与智能手机软键盘相当的输入速度(35WPM¹)和接近100%的准确率。

空中双手盲打输入的运动模型

在三维交互界面上,虚拟键盘十指盲打的文本输入方式可以利用人们在物理键盘上的肌肉记忆,从而有效减少学习成本,同时达到很高的输入效率。

¹ 每秒35个单词。

然而，由于缺少触觉反馈和键盘布局的参考，人们在空中的双手盲打输入信号具有很大的噪声，为输入意图的识别带来了很大的挑战。针对这一交互接口，我们通过一系列的运动模型扩充了贝叶斯算法中点击模型的先验知识，最终实现了较好的文本输入效果。

落点的三维分布模型

在空中打字的场景中，因为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，所以用二维高斯分布来对触点位置进行建模是不够的。同时，由于缺少键盘布局的参考，使得用户落点的空间分布更加分散，难以推断输入的目标字母。为此，我们将触屏上的点击模型进行了推广，根据中心极限法则，针对每个按键，用三维高斯分布来对落点的位置进行建模（见图2）。

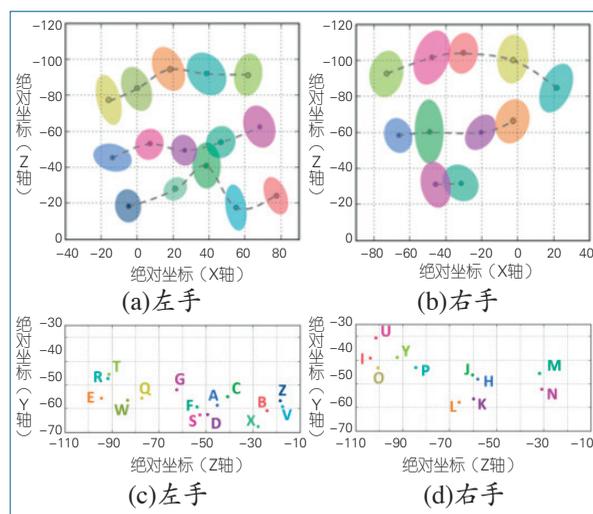


图2 落点的三维分布，分别显示了对应每个按键的落点平均位置和一倍标准差轮廓椭圆

空中多指联动模型

在点击模型中，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，从而难以做出比较准确的区分。针对这个问题，我们利用了十指盲打的一个特点来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和W）往往是由不同的手指点击，如果能利用上手指的信息，将会对落点位置的解释带来极大的帮助。例如，如果Q和W之间的落点是由小拇指触发的，那么用户点击Q的概率更大一些，如果由无名指触发，那么点击W的概率更大。

基于上面的观察，我们希望能在点击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。即在一个手指按下时（如中指），其他手指（如无名指）也会同时向下按。联动现象导致我们观察到的是多个手指同时按下，无法准确区分当前是哪个手指在触发点击。

为了解决这个问题，针对观察到的一次多指点击，我们为每个触发手指分别计算点击的概率。我们首先通过实验观察到一个结果：当主动手指点击时，不同随动手指的点击幅度是不同的，而且具有固定的规律。例如，当无名指点下1cm时，中指平均会同时下移0.51cm，而食指平均会下移0.33cm。据此，我们定义了AR来量化手指之间的联动程度，其数值等于随动手指的点击幅度除以主动手指的点击幅度，并利用一维高斯模型对随动手指的移动幅度进行了建模。测得的多指联动模型如图3所示，从而可以计算单次点击来自各个手指的概率。

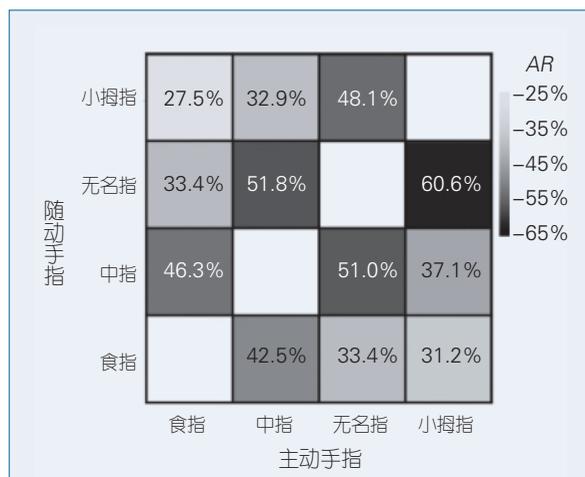


图3 每种主动-随动手指组合的AR

先验知识扩展的贝叶斯模型

综合以上方面,我们利用三维的空间点击模型和空中多指联动模型,建立了空中双手盲打行为的运动模型。相比于文本输入的基本贝叶斯方法中点击模型只考虑落点位置的分布,我们的运动模型显著扩充了用户行为的先验知识,从而能提升贝叶斯方法的推理效果(见图4)。具体而言,我们将公式1进行了改写,引入了每个手指的点击幅度 D ,与之对应,我们需要计算的概率变为 $P(W|I, D)$ 。基于贝叶斯公式,该条件概率正比于 $P(W)P(I|W)P(D|W)$,其中点击模型 $P(I|W)$ 用三维空间触摸模型来计算,而 $P(D|W)$ 描述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。

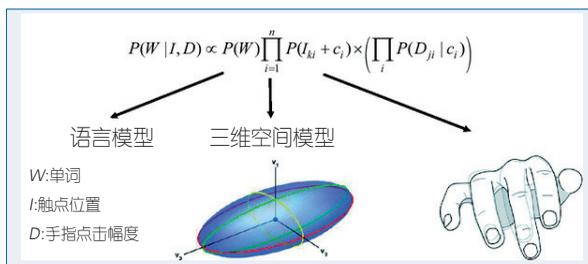


图4 扩展的贝叶斯模型

通过结合语言模型、触摸模型和空中多指联动模型,我们的扩展贝叶斯算法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,从而大幅度提升了输入准确率。用户实验表明,优化的算法首次实现了空中双手盲打的输入体验,可以使用户在接近30WPM的输入速度下,达到近100%的输入准确率,在空中打字这个已存在数十年的交互概念上实现了技术突破。

结语

随着自然人机交互界面的发展,文本输入的方式也发生着日新月异的变化。在多种多样的自然文本输入中,随着交互自然性的提升,输入信噪比低的难题将会越来越凸显。而为了实现高效准确的输入意图推理,文本输入技术亟须通过对人的输入行

为建模,引入更丰富的先验知识,从而突破对数据的依赖,实现兼顾高准确率和可解释性的新一代自然文本输入方法。 ■



喻 纯

CCF 专业会员。清华大学副研究员。主要研究方向为人机交互。

chunyu@tsinghua.edu.cn



易 鑫

CCF 学生会会员。清华大学计算机科学与技术系博士生。主要研究方向为人机交互。

yix15@mails.tsinghua.edu.cn



史元春

CCF 常务理事、会士, CCCF 前副主编。清华大学计算机系“长江学者”特聘教授,人机交互与媒体集成研究所所长,清华大学全球创新学院 GIX 院长。主要研究方向为人机交互、普适计算等。

shiyc@tsinghua.edu.cn

参考文献

- [1] Goodman J, Venolia G, Steury K, et al. Language modeling for soft keyboards[C]//Proceedings of the Eighteenth National Conference on Artificial Intelligence. American Association for Artificial Intelligence, 2002:419-424.
- [2] Findlater L, Wobbrock J. Personalized input: improving ten-finger touchscreen typing through automatic adaptation[C]//Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2012: 815-824.
- [3] Goel M, Jansen A, Mandel T, et al. ContextType: using hand posture information to improve mobile touch screen text entry[C]//Proceedings of the SIGCHI conference on human factors in computing systems. ACM, 2013: 2795-2798.
- [4] Azenkot S, Zhai S. Touch behavior with different postures on soft smartphone keyboards[C]//Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services. ACM, 2012: 251-260.
- [5] Findlater L, Wobbrock J O, Wigdor D. Typing on flat

glass: examining ten-finger expert typing patterns on touch surfaces[C]//Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, 2011: 2453-2462.