

知乎



干货|个性化推荐系统五大研究热点之用户画像（四）



第四范式...
已认证的官方帐号

关注他

28 人赞同了该文章

【编者按】微软亚洲研究院社会计算组的研究员们从深度学习、知识图谱、强化学习、用户画像、可解释性推荐等五个方面，展望了未来推荐系统发展的方向。

在前三篇文章中，我们分别介绍了深度学习、知识图谱、强化学习在推荐系统中的应用以及未来可能的研究方向。在今天的文章中，我们将介绍推荐系统中的用户画像。

推荐系统中的用户画像

构建推荐系统的核心任务之一在于如何准确地分析出用户的兴趣特点，也就是我们常说的用户画像。

简而言之，用户画像是指从用户产生的各种数据中挖掘和抽取用户在不同属性上的标签，如年龄、性别、职业、兴趣爱好等。

赞同 28 1 条评论 分享 喜欢 收藏 申请转载 ...

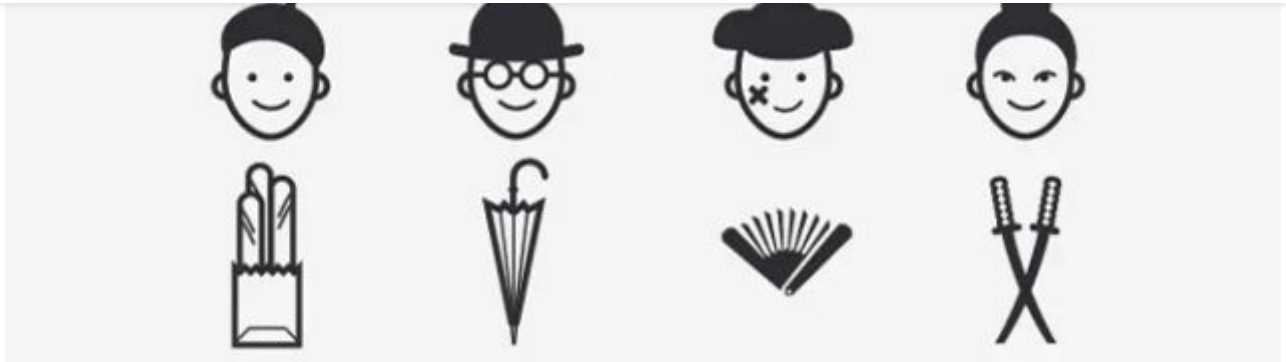
字体越大，特征值越明显

目前，主流用户画像方法一般是基于机器学习尤其是有监督学习的技术。这类方法从用户数据中抽取特征来作为用户的表示向量，并利用有用户属性标签的数据作为有标注数据来训练用户画像预测模型，从而对更多的没有标签的用户的属性进行预测。

首先，这些已有的方法大多数都基于手工抽取的离散特征，这些特征无法刻画用户数据的上下文信息，因此对于用户的表征能力较为有限。

最后，已有的用户画像方法大都没有考虑用户属性标签的时效性，因此很难刻画用户动态变化的属

知乎



从多源异构用户数据中构建深度、统一和动态的用户画像

为了应对上述挑战，我们认为应该从以下方面展开用户画像研究：

1. 构建具有更强表征能力的用户表示模型。

随着深度学习技术的发展和成熟，利用深层神经网络从用户原始数据中自动抽取深层次的、有信息量的特征来构建用户的特征表示能够有助于更加充分地利用用户数据并有效提升用户画像的精度。

使用基于深层神经网络的用户表示模型能够有效克服目前已有的基于特征工程和线性模型的用户画像方法的不足。我们提出的HURA模型(*Neural Demographic Prediction using Search Query*, WSDM 2019)基于多层注意力机制和神经网络结构，有效地通过搜索日志预测了用户个人属性。

2. 基于多源和异构数据的用户画像。用户产生的数据往往分布在不同的平台，并且具有不同的结构（如无结构的社交媒体文本数据和有结构的电商网站购买记录等）和不同的模态（如文本数据和图像数据），给用户画像带来了很大的挑战。

如何设计一个深度信息融合模型来利用不同来源、不同结构和不同模态的用户数据进行用户建模，是未来用户画像领域的一个重要方向。

基于深度神经网络的协同学习和多通道模型可能是值得尝试的技术。

知乎



3. 不同平台用户画像数据的共享和用户隐私保护。目前很多用户数据存在于不同的平台当中，例如搜索引擎拥有用户的搜索和网页浏览记录，电商网站拥有用户的商品浏览、购物、收藏和购买信息。这些不同平台的用户数据对于用户画像都具有重要的价值，互相之间可以提供互补信息，有助于构建更加丰富全面的用户表示。

然而，平台之间直接共享用户信息可能会使得用户的隐私受到泄露和损害。**如何在不转移和不共享用户数据的情况下，充分利用不同平台的用户信息实现协同用户画像和建模是值得研究的一个方向。**

4. 面向用户画像的统一用户表示模型。已有的用户画像方法在实际的应用中往往会涉及大量模型的训练、存储和调用，时间和空间的复杂度都比较高，使用起来也比较繁琐。另外，不同的用户属性之间潜在的联系也无法充分挖掘。

如何基于多源异构的用户数据构建一个统一的用户表示模型，使得该模型可以尽可能全面而准确地包含一个用户在不同属性和维度的特征信息并能够应用于多个用户画像任务是一个非常值得研究的方向。

基于深层神经网络的多任务学习技术和类似词嵌入的用户嵌入技术有希望能够应用于这个问题。

下一篇文章我们将围绕“**推荐系统中的用户画像**”的研究展开讨论。想要了解关于推荐系统的更多研究热点，还请持续关注。

相关阅读：