高端装备制造产业科技服务SaaS用户画像研究

李松江 赵家伟 杨 迪 张 昕
（长春理工大学，吉林长春 130022）

摘要：综合科技服务平台的研发、建设和实施推动科技资源发现系统技术的飞跃性发展。通过大数据技术对海量用户行为数据分析能够帮助系统掌握用户重要特征，构建用户画像，从而提供更加精准的科技服务。本文围绕哈长城市群，面向高端装备制造产业科技服务SaaS应用平台，开展用户画像获取方法研究。通过整合用户静态数据与动态行为数据，完成数据清洗、聚类及标签映射工作，进而提取用户特征，在高端装备制造产业科技服务中实现精准营销、个性化服务，强化整体科技服务体系，使科技服务更加方便、快捷。

关键字：用户画像；用户需求；行为数据；科技服务；用户特征
中图分类号：TP520 **文献标识码：**A **DOI：**10.3772/j.issn.1674-1544.2020.05.003

Research on SaaS User Portrait of High-end Equipment Manufacturing Industry Technology Service

LI Songjiang, ZHAO Jiawei, YANG Di, ZHANG Xin
(Changchun University of Science and Technology, Changchun 130022)

Abstract: The research, development, construction and implementation of integrated science and technology service platform promote the rapid development of science and technology resource discovery system technology. The analysis of massive user behavior data through big data technology can help the system master the important characteristics of users, realize the construction of user profiles, and provide more accurate scientific and technical services. This paper focuses on Harbin-Changchun urban agglomeration, and the SaaS application platform of high-end equipment manufacturing industry technology service, and carries out the research on user portrait acquisition methods. Through the integration of user static data and dynamic behavior data, data cleaning, clustering and label mapping are completed, and then user characteristics are extracted, so as to achieve precise marketing and personalized services in high-end equipment manufacturing industry science and technology services, strengthen the overall science and technology service system, and make science and technology services faster and more convenient.

Keywords: user portrait, user demand, behavioral data, scientific and technical services, user characteristics

作者简介：李松江（1984—），男，长春理工大学研究生导师，主要研究方向：数据挖掘；**赵家伟**（1996—），男，长春理工大学硕士研究生，主要研究方向：大数据；**杨迪**（1990—），女，长春理工大学讲师，主要方向：智能交通（通信作者）；**张昕**（1974—），男，长春理工大学讲师，主要研究方向：计算机网络。

基金项目：国家重点研发计划现代服务业重点专项项目“哈长城市群综合科技服务平台研发与应用示范”（2017YFB1401800）。
收稿时间：2020 年 3 月 23 日。

0　引言

高端装备制造业是我国装备制造业的高端领域，是典型的知识密集、技术密集和多学科交叉集成的多科技产业，对国家战略性经济建设与产业升级起到举足轻重的作用[1]。我国高端装备制造产业资源丰富，但资源分散，产业间的资源整合困难。通过构建高端装备制造业科技服务平台，可以整合科技服务资源，实现资源共享、重用，避免资源浪费。而个性化服务是高端装备产业科技服务平台的重要组成部分。用户画像在个性化服务中有着重要的作用。词汇“用户画像”早期应用于美术界，但随着进入数据爆炸的新时代，“用户画像”则被定义为通过用户在日常互联网生活行为下产生一系列数据精炼出来的特征标识[2]。构建以用户画像为基础的个性化服务能够满足新时期用户对服务精确性、及时性、专业化的需求，大大提升科技服务质量，有助于科技服务更高效地整合服务资源完成供需方需求匹配工作，促进整个科技服务新技术新产业的萌发。哈长城市群是东北经济和科技创新的新引擎和新高地，打造一体化科技服务平台对于整合科技服务资源，带动地区产业创新、技术创新，进一步提高东北老工业地区高端制造业的整体水平具有重要的意义。哈长城市群高端装备制造科技服务平台利用SaaS灵活、易扩展、成本低的服务特性构建了SaaS系统，提升了SaaS用户的使用体验。本文将围绕哈长城市群，面向高端装备制造产业科技服务SaaS平台，从多个方面讨论用户画像在科技服务中的重要作用，分析用户行为特征，并根据供需方多元历史数据分析构建用户画像，为不同用户提供个性化高端装备制造产业科技服务。

1　用户画像需求分析

1.1　对市场供需方提供需求分析

通过对用户画像的研究可以提升科技服务质量、完善科技服务流程、强化科技服务体系，使得科技服务更加方便、快捷。对于市场供需方来说，用户画像的获取可以极大地节约业务流程时间，缩短工期，减少因为筛选彼此所造成的时间资源浪费，增强科技服务便利性。科技服务业务流程如图1所示。供需双方在服务平台上发布（或接受）任务，并进行业务交流。若仅将需求方所要发布的任务发布在平台上，则会造成大量数据的堆叠。由于显示界面篇幅小，新发布的需求任务会覆盖过去的需求任务，使供应方难以及时收到之前发布的需求任务信息。在整个任务流程中，至关重要的是供应方如何在大量需求任务中去选择最适合自己的需求任务，需求方如何在

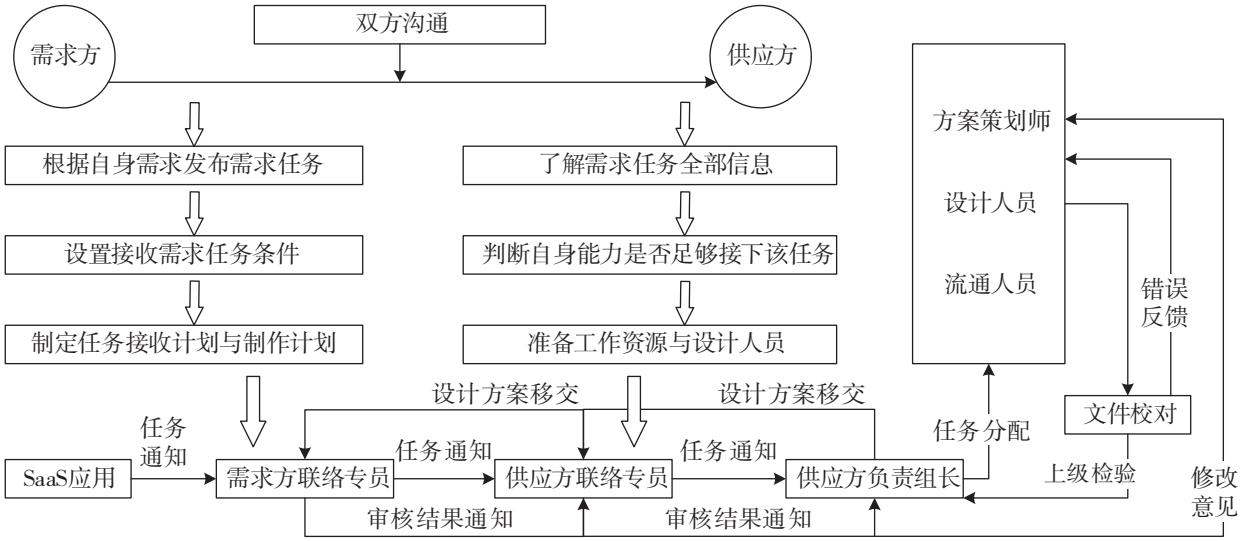

图1　科技服务流程图

大量申请者中去选择最优质的供应方。

高端装备制造产业所涉及行业广泛，不仅包括小型制造业，如高端精密仪器零件的制作，而且包括大型制造业，如汽车、小型卫星的制作与组装。用户画像的获取加强了供需双方与上述各类行业需求任务的“紧密度”，让整个业务流程更加清晰，让整个科技服务流程更加便利。根据供需方用户大量的数据行为进行分析，得出用户特征，构建他们的用户画像。系统根据获取的用户画像进行下一步的操作，如通过个性化服务将更适合该供应方的需求任务在科技服务平台上直接显示给供应方、根据供需双方产生的用户画像揭示二者之间的关系、根据画像中用户的任务完成度等因素推断企业下一步的发展方向等，从而帮助供应方在大量的需求任务中挑选出最适合自己的需求任务。和供应方一样，需求方也在“筛选”中充满了困扰。如在其发布了需求任务就会得到供应方的申请，若申请的供应方数目过多，需求方则需要花费更多的时间筛选供应方。因此，对供应方和需求方的相关信息进行用户画像，可以方便需求方根据供应方的用户画像迅速筛选出合适的供应方，并通过分析供应方的其他信息，如历史合作供应方、任务发布完成度等，可以更加顺畅地对申请者进行筛选。从而帮助需求方在大量的申请者中选择最优质的供应方。

1.2 对科技服务提供者提供需求分析

用户画像的获取让整个SaaS应用供需方的操作变得更加便利，极大地提高了科技服务的便利性，同时对于整个科技服务的提供者也起着参考作用。科技服务提供者根据用户画像进一步完善了整个科技服务体系，不断提高科技服务的创新性，增强了服务科技含量。根据用户画像整合的数据挖掘用户偏好，使供需方用户在完成申请任务、筛选申请者、接收任务信息时感到愈发便捷。此时，科技服务的辐射带动性就很好地体现出来，用户会有意识地发觉整个平台在不断进步，以单个用户为节点进行扩散，增加平台用户数目，扩大整个平台规模。科技服务提供者通过观测用户画像信息确定整个科技服务平台发展的方向，保持在同行业服务水准的领先[3]。不断观测大量用户画像会使科技服务提供者增强其业务逻辑能力及分析能力。在数据量、用户画像数目不断增多的情况下，科技服务提供者可以进一步增强后续算法，既有助于自身学识的不断进步，又有助于平台的发展。

1.3 对政府管理者提供需求分析

用户画像的获取让整个科技服务平台SaaS应用成为高端装备制造业和哈长城市群的纽带。在过去，想要了解城市群之间的信息必须经过多方媒介进行传递，且有大量用处不大的信息干扰着政府管理者的判断。而如今，用户画像的出现，可以及时地将有用的信息传递到整个城市群。政府管理者可以根据这些信息更好地把握整个行业的发展方向，引领整个行业的走向并根据这些信息制定行业政策，促进行业的发展及至整个城市群的发展。

2 用户数据来源整合

用户画像的获取离不开大量的数据支持，在整个科技服务平台所形成的静态数据与用户行为就是数据源的主要组成部分。数据源可以细化为用户本身的静态数据、用户在SaaS应用操作过程中形成的历史数据、每次任务流程中形成的日志数据、任务每个流程的完成情况等。用户画像构建的核心就是用户在这些基础数据和行为下所显现的用户特征[4]。这些特征既包括基础数据下用户的静态数据特征，也包括用户在与不同用户进行交流之后的动态数据特征。因此，在获取用户的基础数据的同时还要获取用户交互数据、基础资源数据。面对大规模异构数据需要构造采集框架以获取分析所需的特定数据[5]。

用户数据从用户所具有的资产情况、人才数量、任务完成数目、公司主营类型、任务发布数目等信息中获得。这些数据在一定程度上代表了供应方的规模和实力，可以帮助需求方在审核供应方时判断是否能够完成需求方的任务。供应方的任务完成数量也是一个重要的数

据来源。如一个公司仅完成了几项任务，且公司实力不是十分雄厚，那么对于那些大规模的需求任务，大多数的需求方在筛选时可能就会将这些公司自动过滤掉。而任务完成情况在考察供应方时也是相当重要的。当一家公司接了许多任务，且几乎每一项任务都按期完成交割，就一定会给许多需求方留下好的印象。当一家公司对于许多任务的完成度并不高，还出现产品质量及设计方案频频返工等问题，这家公司的画像在任务完成度、客户评价满意度等特征方面一定是很低的。需求方在筛选众多候选申请者时就不会考虑这样的供应方。

对需求方来说，在平台上发布的任务数目越多，其公信力就越强，就会有更多的供应方愿意接下其发布的需求任务。一家需求方所发布需求任务的类型一般都是相似的。如果该需求方发布了和过去任务类型相似的需求任务，说明需求方对该类型需求任务的整体业务流程非常熟悉，那么和供应方进行任务交接时就会非常顺利。如果其发布了和过去任务类型完全不同的需求任务，那么就需要重新了解很多的业务流程，可能在项目审核与交接过程中给供应方带来不便与困扰，甚至耽误整个工作的进度。在供应方去申请需求任务时也会参考该需求方发布的全部历史任务，若该需求方在其发布的很多需求任务上都存在着多次审核等情况，那么供应方就会进一步考量该需求方是否有着过分严格的要求。

3　用户画像构建

构建用户画像的整体流程如图2所示。首先进行SaaS应用上用户数据采集。其次根据每个任务流程上所产生的动态数据对供应方的整体任务完成能力进行评估，获取所有需求任务的主题进而获得需求方（供应方）发布（接受）主题任务的偏好。然后采用多维度标注法对用户进行分类，展示用户与用户之间、供应方与需求方之间的关系。最后构建用户画像。

3.1　数据采集

用户画像本质是用户特征的“可视化”。用户画像分为静态数据画像及动态数据画像两个部

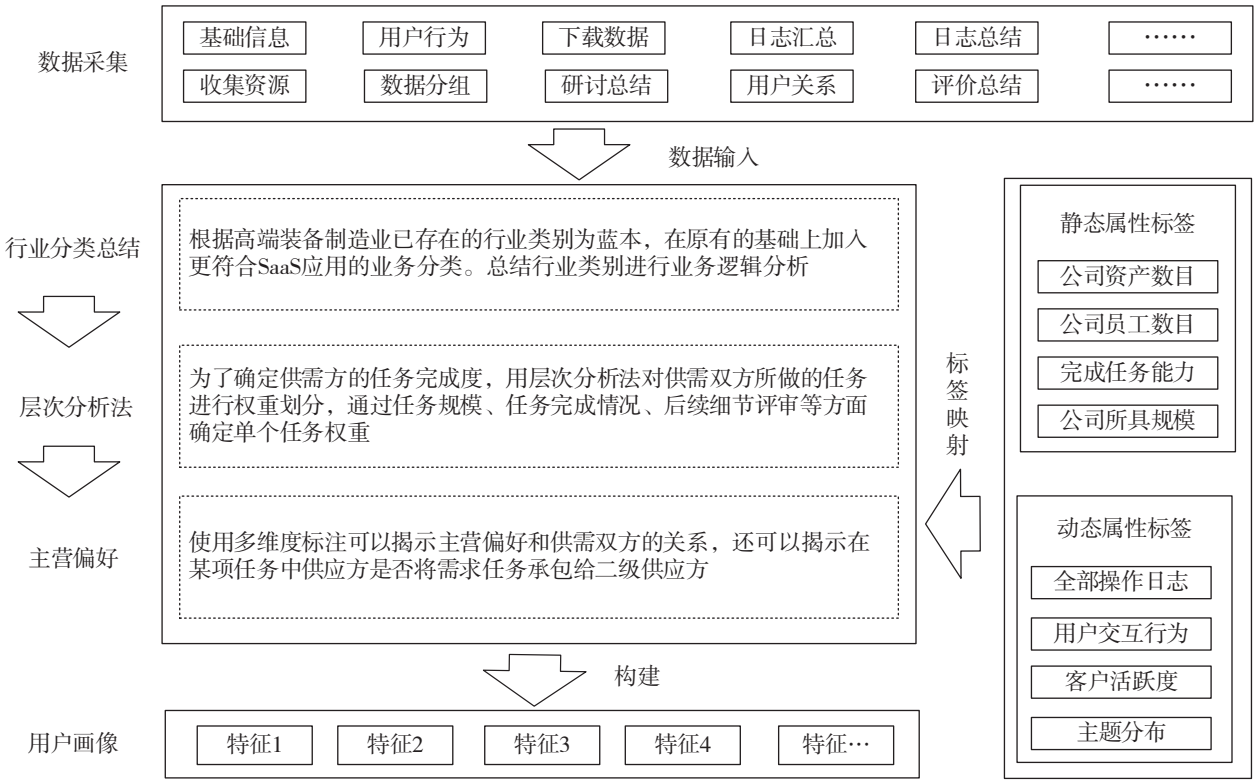

图2　基于高端装备制造产业科技服务的SaaS用户画像构建流程图

分。静态数据画像是根据用户所具有的固有属性数据得到；动态数据画像是根据行为驱动的动态需求特征得到，可以从用户的各种行为数据中挖掘，按照对于不同行为的权重划分构成用户画像的各类标签[6]。

用户画像获取首先要收集用户操作形成的大量历史数据，这些操作主要包括在整个SaaS平台上进行的业务起始操作、业务流程操作、公司内部操作（图3）。其中用户交互行为还涉及需求方对于供应方所提交任务计划书、设计方案的审核。对采集的大量历史数据进行分析与处理，可以将本来杂乱无章的数据变成有用的数据。

根据需求方所发布需求任务的类别进行需求任务分类，形成任务类别体系。以用户ID作为整个数据获取的根节点，获取标准的用户画像。用户所具有的资产、人员、公司规模等属于静态用户数据。用户社交数据、任务操作历史数据等构成用户动态画像数据。最后对在科技服务平台上所得到的数据进行数据清洗、分类、聚类等[7]。

3.2 任务完成情况

在用户画像获取过程中，最重要的是了解一个公司的任务完成情况。一个需求任务从发布到评价都离不开参考公司的任务完成情况，其在整个业务流程中起着十分关键的作用。在需求方发布需求任务后，供应方要去申请需求任务。供应方通过查看该需求方完成任务数目、同类别需求任务完成数目、资金回拨情况评估自身是否要接下该需求任务。SaaS应用将需求任务个性化推荐给供应方时，除了考虑供应方主营任务类型，还要考虑该供应方的任务完成情况。需求方在审核供应方时，任务完成情况是其考察的一个重要因素。任务完成情况不仅代表了供应方对于各类任务的熟练度，而且代表了供应方面对需求任务的态度。需求方根据上述因素判断供应方能否有条件接受该需求任务。

用户总体任务完成情况由全部单个任务的完成情况构成。对于供应方而言，单个任务的完成情况从以下几个方面进行评估。一是是否超时：从接受该需求任务开始截止到完成任务这段时间是否满足需求方所规定的时间。二是退回次数：供应方在设计完成任务图纸之后提交给需求方，需求方对图纸进行审核，如果不合格则被需求方退回。三是评价情况：当供应方完成该需求任务后，需求方从任务开始到整个任务完成的所有业务流程进行综合评价。

将一名用户所有任务完成情况整合起来，则需要用户所做的每个任务的完成情况。根据层次分析法得出决定单个任务完成情况所占总体任务完成情况的因素权重（图4）。得出权重后各任务按因素加权合并得出该用户完成任务的能力[8]。

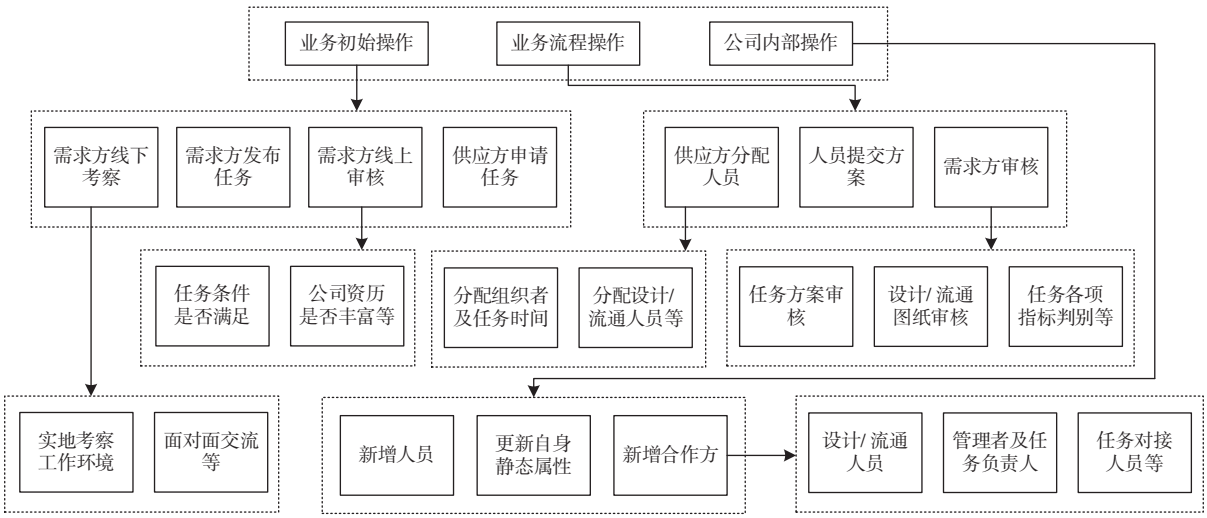

图3 用户操作行为

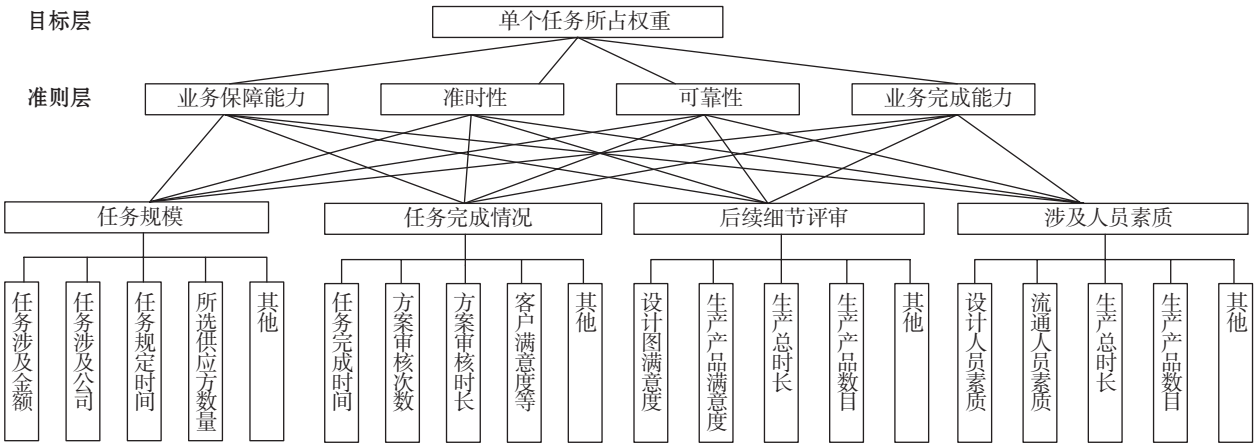

图 4　层次分析法确定权重

3.3　主营偏好获取

根据用户所发布（或者接受）的任务类别以及任务数量，可以发现该用户的主营偏好。对于供应方而言，深入挖掘数据文本能够看到该用户对高端装备制造行业的哪一类任务更感兴趣[9]，对哪一类任务的完成度较高。对于需求方而言，发布的需求任务类别往往也是固定的，从数据直观显示上就可以确定其主营类别。

在整个科技服务平台上还存在部分二级供应方。这部分二级供应方由于自身公司规模的局限性往往不能直接接受需求方发布的任务，这时他们就会依赖一级供应方从需求方处接受任务。因此，为了确定需求任务是否由一级供应方直接完成，在获取用户画像时还需标注是由供应方本身直接完成需求任务还是由二级供应方完成。

3.4　提取用户特征构建用户画像模型

在上文得出用户任务完成情况、用户的主营偏好等用户特征的基础上，使用多维标度法揭示供需双方的关系，构建用户画像。根据该方法可以实现整个研究主题中每个多维对象的相对空间位置。相对位置如果越接近对象，则与对象之间的相似程度就越高。依此可以基于主题对用户进行分类。多维标度法可以直观地显示同一需求方发布的不同需求任务特点[10]，从而进一步研究当需求方发布任务时应该将其推荐给什么样的供应方去完成。对整个供应链进行区分可以达到优化流程效率的作用。该方法与其他常用的分类方法如决策树分析法、关联规则等相比较，其优势更加明显。

在整个用户数据中，既有用户静态数据，又有大量的动态数据。在特征如此多的情况下需要解决多变量因素影响结果的问题。所谓的因变量是更为复杂的群体类型，对此更要筛选出不同用户的特殊取值，提取出不同用户的典型特征[11]。

在提取典型特征后，就构成了用户画像。画像内容如图 5 所示，主要包括合作关系、基本特征、公司业务能力、行为特征、心理特征等方面。其中，合作关系包括与之合作的全部公司、与公司合作的次数等；基本特征包括公司成立时间、公司资产数目、公司员工数目等；公司业务能力包括接受（发布）的任务数目、所接受（发布）的任务规模、完成任务的情况等；行为特征包括同时申请大量需求任务、经常找二级供应商承包等；心理特征包括接受（发布）的任务类型偏好、合作公司类型偏好、内心抵触任务拖延、抵触二级供应商承包等。

4　结语

用户画像的研究对于高端装备制造产业科技服务提供了很大的便利性，大大节省了业务流程时间，提高了科技服务质量。对于政府管理者而言，画像的分析便于其了解整体行业的发展趋势，加强供需方之间的联系。同时，对于整个哈长城市群而言，用户画像的构建打破

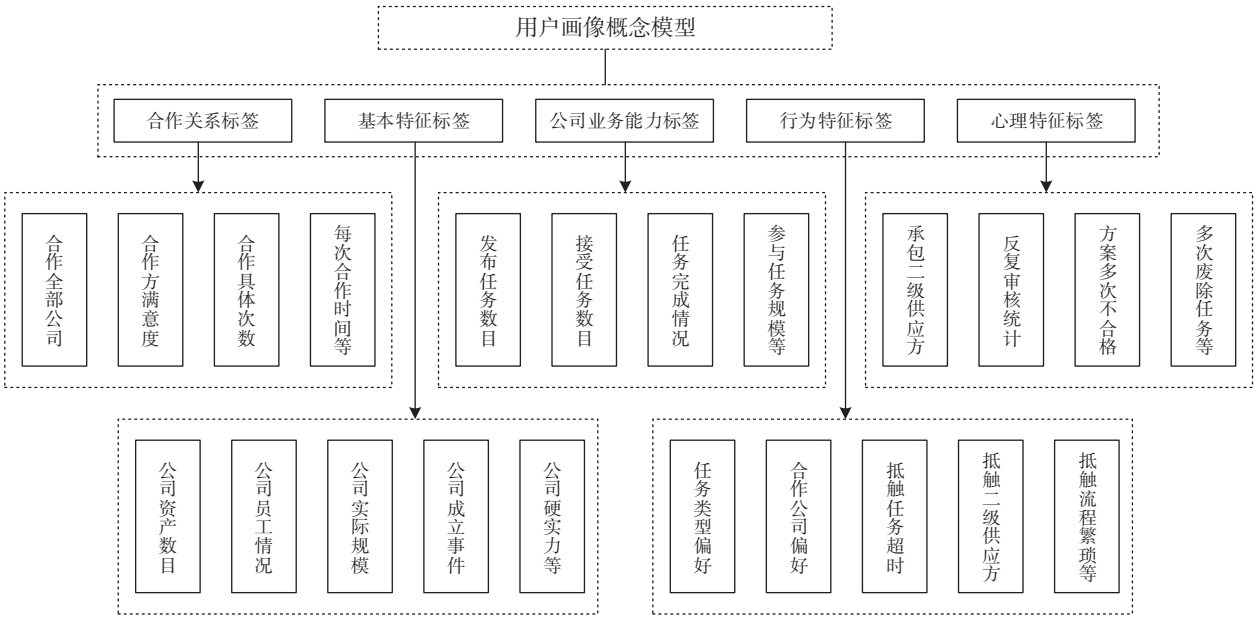

图 5　用户画像概念模型

了地域企业间信息共享互通的壁垒，深化了企业间的联系，加强了企业间的合作。由于诸多主客观因素存在，本研究也存在着局限性。一是数据量问题。整个研究是基于平台展开的，而由于科技服务平台推出的时间较短，产生的用户数据量较少，在画像构建和特征提取等方法上难免存在误差。二是异构数据融合问题。随着数据量增多未来会出现多源异构数据融合问题，这也是今后用户画像研究的重点与难点。用户画像具有较强的实践指导性，本文提出的研究参考框架主要基于哈长城市群高端装配制造业SaaS应用平台的实践，今后可以从更多行业、企业实践中选取典型案例，针对本文研究的局限性继续完善用户画像研究。

参考文献

[1] 肖小平, 周光辉, 张超, 等. 面向高端装备产品设计的多模式知识服务方法[J]. 应用科技, 2019, 46(1): 11−18.

[2] 亓丛, 吴俊. 用户画像概念溯源与应用场景研究[J]. 重庆交通大学学报(社会科学版), 2017, 17(5): 82−87.

[3] 丁晓东. 用户画像、个性化推荐与个人信息保护[J]. 环球法律评论, 2019, 41(5): 82−96.

[4] 费鹏, 林鸿飞, 杨亮, 等. 一种用于构建用户画像的多视角融合框架[J]. 计算机科学, 2018, 45(1): 179−182, 204.

[5] 丁伟, 王题, 刘新海, 等. 基于大数据技术的手机用户画像与征信研究[J]. 邮电设计技术, 2016(3): 64−69.

[6] 林燕霞, 谢湘生. 基于社会认同理论的微博群体用户画像[J]. 情报理论与实践, 2018, 41(3): 142−148.

[7] 杨颖琦, 郑南章. “大数据画像” 构建人才培养体系[J]. 中国电力企业管理, 2019(3): 77−78.

[8] 周俊杰, 赵晓萌, 方少亮, 等. 科技资源共享服务中用户“画像”研究[J]. 中国科技资源导刊, 2019, 51(4): 59−68, 77.

[9] 黄文彬, 徐山川, 吴家辉, 等. 移动用户画像构建研究[J]. 现代情报, 2016, 36(10): 54−61.

[10] 欧阳秀平, 廖娟, 冯烨, 等. 基于运营商大数据的游戏用户画像构建研究[J]. 邮电设计技术, 2019(9): 40−44.

[11] 王显斌. 基于用户画像的科学数据管理模型研究[J]. 电子技术与软件工程, 2019(17): 179−180.