

基于用户行为数据的人格测量

张富峥 谢幸
微软亚洲研究院

关键词：人工智能 人格测量 大五人格

心理学中的人格理论

善于驾驭文字的作家寥寥数笔就能生动地勾勒出笔下人物的特征，好的文学作品对人物的描述绝不仅仅限于他的容貌和行为，更重要的（至少是同等重要的）在于对人物性格的描绘和刻画。在心理学领域，这便是对人格（personality）的阐述和展示。

虽然“人格”这个术语在日常生活中很常见，但是给人格下一个准确清晰的定义却并非易事，即使是心理学家在这个术语的定义上也很难达成共识。最早的对人格的定义可以追溯到公元前400年古希腊医学家希波克拉底（Hippocrates）的体液说。他认为，人体是由四种体液构成的，即血液、粘液、黄胆汁和黑胆汁。而人格可以被定义为四种类型，并且与体内的这四种体液息息相关：黑色的胆汁产生了忧郁型人格，红色的血液产生了乐观型人格，黄色的胆汁产生了易怒型人格，粘液产生了冷静型人格。尽管希波克拉底的体液说已经被现代医学所否定，但是他关于人格分类的探讨，对于心理学中这一术语的定义仍然是有启发意义的。

在现代心理学中，人格的定义其实和语言的使用有着紧密的关系。每当我们试图对人格特征做出描述时，我们总会去寻找一些心中自认为恰如其分的词语。比如，描绘一个人物是否是冷静的、客观的、负责任的，还是相反。在日常生活中，我们似乎更倾向于使用高度概括而又简洁的方式去描述人格（借助已经发展成熟的语言系统，使用规范化的词汇），而不是使用像“他在如此险恶的环境下仍然保持着好心情”这些间接的、

仍然需要进行认知加工理解的语句。这些简洁的词汇描述了种种相同、相似或者相异的人格特质。基于这些观察，人格理论的先驱奥尔波特（Allport）和奥德伯特（Odbert），在1936年对英语词汇进行了艰难而又系统的调查研究，并按照个人特质、暂时的情绪或者行为，以及智力与才干这四个类别，列出了大约18000个单词，其中大约有5000个“可能存在的用于描绘个体特质的中性词汇”。他们的研究得到了众多心理学家的特别关注^[1]。

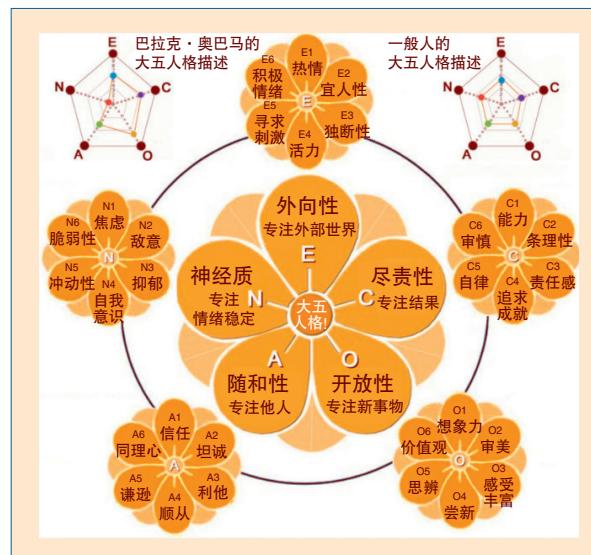
然而，用这些成千上万的词汇来定义人格，对于心理学家可以说是一场灾难，特别是在那个没有计算机的时代，被他们称之为“like a semantic nightmare”^[2]。为了能够系统地区分和比较不同人的人格特质，对这些成千上万的词汇进行归类，以便建立人格的分类法（taxonomy）就显得非常必要。因此，心理学家开始采用因素分析（factor analysis）^[3]技术对奥尔波特及奥德伯特的单词表进行分析，从而得到了数量较少的相关单词群组，每个单词群组可以认为是人格某一方面的特征。因素分析是一种统计方法，根据大部分人对一系列题目的回答之间的相关性，将相关的项整合为“因素集合”，其中每个因素可以认为是与之相关的所有词汇的高度概括。例如，某种人格定义的某一因素可能与外向性有关，这个外向性因素就是与之高度相关的所有词汇的概括。最具有影响力的因素分析理论家是卡特尔（Cattell）。在20世纪60年代中期，他精心挑选出奥尔波特和奥德伯特单词表中的200个单词，要求人们对他们的朋友和自己进行评定，并把这些评定结果进行因

素分析，最后得出了著名的 16 种人格因素量表 (16 personality factors)^[4]。表 1 展示了一个人的人格特征可以被描述为 16 种因素的得分情况。

表1 卡特尔16种人格因素¹

因素	低分描述	高分描述
A. 乐群性	缄默，孤独，冷漠	外向，热情，乐群
B. 聪慧性	思想迟钝，抽象思考能力弱	善于抽象思考，学习能力强
C. 稳定性	情绪激动，易受环境支配	情绪稳定而成熟，能面对现实
E. 恃强性	谦逊，顺从，通融，恭顺	好强固执，独立积极
F. 兴奋性	严肃，审慎，冷静，寡言	轻松兴奋，随遇而安
G. 有恒性	苟且敷衍，缺乏奉公守法的精神	有恒负责，做事尽职
H. 敢为性	畏怯退缩缺乏自信心	冒险敢为，少有顾忌
I. 敏感性	理智的，着重现实，自恃其力	敏感，感情用事
L. 怀疑性	依赖随和，易与人相处	怀疑，刚愎，固执己见
M. 幻想性	现实，合乎成规，力求妥善合理	幻想的，狂放不羁
N. 世故性	坦白，直率，天真	精明能干，世故
O. 忧虑性	安详，沉着，有自信心	忧虑抑郁，烦恼自扰
Q1. 实验性	保守的，尊重传统观念	自由的，不拘泥于现实
Q2. 独立性	依赖，随群附众	自立自强，当机立断
Q3. 自律性	矛盾冲突，不顾大体	知己知彼，自律谨严
Q4. 实验性	心平气和，闲散宁静	紧张困扰，激动挣扎

然而，卡特尔的 16 种人格因素仍然受到了广泛的质疑。一些人格研究者无法重复验证 16 个独立的因素，而且许多人提出，采用较少的因素就可以捕捉到个体差异的重要方面。近 20 年来，人格研究者关注与支持最多的人格定义是五因素模型，也常常被称为“大五人格”。大五人格的定义如图 1 所示，包括五个高度概括的人格因素：外向性 (extraversion)、尽责性 (conscientiousness)、神经质 (neuroti-

图1 大五人格²

cism)、随和性 (agreeableness) 和开放性 (openness)，以及每个人格因素下的一些细分特质（比如外向性下包括了是否经常参加活动、是否热心肠等）。费斯克 (Fiske) 最先发现，在不同的评估方法中，如在自我评定、同伴评定以及观察者评定的评测方法中，都可以得到定义相近的五个因素^[5]。戈德堡 (Goldberg) 也在一系列的研究中发现，即使样本数目不同，问题设计方式以及因素数目选取方式不同，也能提取出一致性相当高的五个因素^[6]。在以使用英语词汇为对象的研究中，五因素模型确实表现出了惊人的可重复验证性，至少有十多名研究者通过不同的样本发现了这个因素。在过去的半个世纪里，五因素结构每十年就会被重复验证一次，说明了它具有跨时间的稳定性，而且在不同的语言环境中也得到了充分的验证。大五人格的建立，使得人格理论研究领域得到了统一。

心理学中的人格测量

¹ 参见<http://baike.baidu.com/view/5104003.htm?fromtitle=%E5%8D%A1%E7%89%B9%E5%B0%9416PF&fromid=8144165&type=syn>。

² 参见<http://www.handresearch.com/diagnostics/extraversion-hand-chart-extraverts-ambiverts-introverts.htm>。

在心理学研究中，人格测量的主流方法包括访谈和调查问卷。访谈需要经过严格训练的人格心理学研究者与个体进行对话，在个体积极配合的情况下获取个体的人格信息。访谈法能够灵活地照顾到个体的差异，并且能够应对特殊情形。但是，访谈法难以广泛使用，因为它必须依赖优秀的专业测评人员才能进行。调查问卷方法需要个体根据自己的实际情况回答问卷上的问题，然后根据答案的分数对应到相应的人格特质中。调查问卷的优点是易于管理和应用，容易推广，答案分数的计算方法客观，并且可以直接获取检测对象的回答。但是调查问卷的局限性也很明显，测验对象可能对自身并没有深入的了解，在回答问题时他们可能会尝试展示自己最好的一面，也可能会尝试根据自己的期望和愿望作答，所以很难得知他们在完成调查问卷时是否精神集中和足够谨慎。尽管存在主观性等局限，调查问卷仍然是心理学目前普遍采用的方法。

以大五人格为例，研究者相继编制出了许多不同类型的调查问卷。其中一些问卷经过研究者的不断验证，具有良好的信效度和适用性，能够广泛应用于实际生活中，比如人力资源管理等场景。根据问卷题目呈现的形式，可以将这些问卷分为语句自陈式、词汇核检表式和图片非言语式。在自陈式问卷中，每个题目均采用了自我描述性的句子或者短语来呈现。比如，“我认为我自己喜欢与他人合作”。NEO-PI^[7]就是最早编制的自陈式大五人格问卷，总共包含了181道题目。由于该测验题目量多，测试所需时间过长，因此科斯塔(Costa)和麦克雷(McCrae)在确保尽量覆盖各个维度和信效度的前提下，编写了只有60道题目的简版NEO-FFI^[8]。词汇核检表式问卷的每道题目由两个意义相反的形容词组成，受测者需要根据自己的情况在两个形容词中直接选择一个分数。比如，在测试随和性维度的题目中，一组形容词的一极是“友好”，另一极是“冷淡”，受测者需要在两极之间的5点量表上打分。比较典型的词汇核检表式问卷有戈德堡提出的TDA^[9]和霍华德(Howard)与麦地那(Medina)提出的BFLQ^[10]。为了消除语言文化因素在人格测验中的影响，研究

者也在努力构建非言语式的大五人格测验问卷。比如FF-NPQ^[11]就是一种图片式的问卷，问卷中的每道题目就是一张图片，图片往往描绘的是某个人物在一个具体情境中表现出某种与其人格特质相关的特定行为。测试中，要求受测者在观看图片时，假定自己是这个图片里的中心人物，然后对自己做出该人物行为的可能性从“非常不可能”到“非常可能”进行评定。表2^[12]总结了不同类型的调查问卷。

表2 针对大五人格的不同版本的调查问卷

类型	名称(开发时间)	编制者
自陈式	NEO PI-R (1992)	科斯塔和麦克雷
	NEO-FFI (1992)	科斯塔和麦克雷
	NEO-FFI-R (2004)	麦克雷和科斯塔
	NEO-PI-3 (2005)	麦克雷等
	NEO-FFI-3 (2007)	麦克雷和科斯塔
	BFI-44 (1991)	约翰(John)等
	BFI-10 (2007)	拉姆斯蒂德(Rammstedt)和约翰
	FFPI (1999)	亨德里克斯(Hendriks)等
	IPIP-BFFM (2001)	戈德堡
	Mini-IPIP (2006)	唐纳伦(Donnellan)等
核检表式	TIPI (2003)	戈斯林(Gosling)等
	SIMP (2005)	伍兹(Woods)和汉普森(Hampson)
	CBF-PI (2010)	王孟成、戴晓阳等
	TDA (1992)	戈德堡
	Mini-Markers (1994)	索西尔(Saucier)
非言语式	LASR-B5 (1990)	特拉普内尔(Trapnell)和威金斯(Wiggins)
	23BBF (1995)	杜依森斯(Duijssens)和迪杰克斯特拉(Diekstra)
	SACBIF (1996)	帕鲁吉尼(Perugini)和莱昂内(Leone)
	BBM-FFM (1999)	谢弗(Shafer)
非言语式	FFM-RF (2006)	穆林斯-史威特(Mullins-Sweatt)
	FF-NPQ (1998)	鲍诺内(Paunonen)等

这些不同版本的问卷，针对信效度、测试时间和文化差异都有不同的偏重，研究时可以根据需要选择适用的版本。

人工智能走进人格测量

在人格测量中，传统心理学中的访谈和调查问卷方法都需要耗费大量的人力、财力和时间，受测者往往局限于几十人到几百人的规模，不可能实现大规模用户的测量。其实，心理学中还有一种人格测量的方法叫做行为测量，它通过观察个体的行为来进行测评。行为测量的理论基础是人格理论中的人类行为的一致性。既然人格能够解释人际之间稳定的个体差异，那么个体行为表现出的差异性就和个体的人格息息相关，因此，使得通过观察个体行为预测人格成为可能。只是在计算机技术得到广泛应用之前，心理学家很难收集到用户足够丰富的行为数据，所以数据的匮乏导致了行为测量在传统心理学中并没有被广泛采用。

随着互联网、智能手机和各种传感设备的普及，用户的行为数据被广泛收集，再加上人工智能方法在建模用户方面的推进，使得通过行为数据测量人格的方法在计算机和心理学的交叉领域得到快速发展。用基于行为数据的人工智能方法进行人格预测，首先需要收集少量用户的调查

问卷结果作为标注，然后通过用户的行为数据提取相关特征，并使用机器学习、深度学习等技术建立

用户特征和人格标注的关联模型，最后使用建立的关联模型根据用户特征进行人格特质的预测。剑桥大学的科辛斯基(Kosinski)和史迪威(Stillwell)等人^[13]在这个领域进行了研究并已取得开创性的成果。他们的结果展示了用人工智能方法对行为数据建模，可预测大五人格的能力。这项研究共收集了近10万名志愿者的信息。研究者要求志愿者填写大五人格的调查问卷，以获得他们的人格特质。为了了解家人和朋友对每个志愿者性格的判断，研究人员还邀请他们在脸书(Facebook)上的好友和家人，通过填写调查问卷来完成。而人工智能做出的个性判断，是根据这些志愿者在脸书上的点赞(like)行为，使用奇异值分解得到用户特征，以他们自我评定的大五人格作为标注数据，然后用机器学习算法来进行建模预测的。通过将参与者的自我评定与人工智能及他人的判断进行比较，研究人员首先检测了自我评定和他人判断的一致性。由于他评受到人际关系的影响较大，因此研究人员又进一步将“他人”细

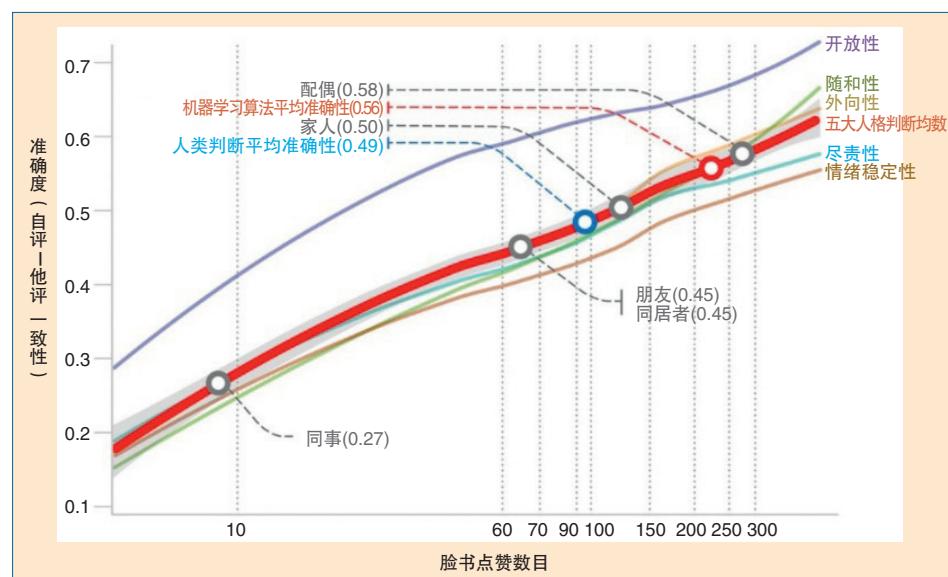


图2 点赞数目和大五人格预测的准确率³

分为朋友、配偶、家人、同居者及同事等几类。

图2阐明了人格判断的准确性与点赞数目之间

³ 参见<http://blog.lavector.com/aid/>。



图3 和大五人格强相关的头像照片

的关系。红线显示的是人工智能对大五人格判断的平均准确性，他人判断的准确性也在这条红线上标出。由于该样本中个体的平均点赞数为 227，由此得到人工智能方法在这个点赞数目下的平均准确性为 0.56，接近配偶判断的准确性 (0.58)，并且远高于家人 (0.5) 和其他好友 (0.49)。研究表明，随着点赞数目的增加，人工智能方法的准确性也有所增强。当点赞数超过 500 时，该判断的准确性可达 0.66。这个结果非常有意思：人工智能只须利用人们在社交媒体上的“点赞”行为，就可以相当准确地对一个人的性格特点进行预测。只需要 10 个赞，人工智能对你的了解就会超过你的同事；70 个赞，超过朋友或者室友；150 个赞，已经超过了你的家人；300 个赞，就连你的配偶，也就是最熟悉你的人，也自叹弗如。

除了脸书上的行为数据，基于国内的社交媒体，研究者们也取得了类似的进展。比如中国科学院心理研究所研究员朱廷劭等人把大五人格的各个维度划分为低分组、中等组和高分组三个类别，通过使用人人网中用户的行为数据（基本信息、情感表达和社交状况等）来预测大五人格的分类^[14]，并采用决策树模型对大五人格进行分类预测，准确率达到 0.7 以上。在另一项研究中，他们使用深度学习中的自动编码技术 (auto-encoder)，学习用户在新浪微博中的语言使用特征，使得大五人格预测的准确性得到了进一步的提升^[15]。

之前的工作都侧重于利用单一领域的行为数据

预测大五人格，我们的研究工作在此基础上更进一步，利用社交媒体上的异构数据（比如头像照片、发表的文字、表情符使用以及社交关系等）来预测大五人格⁴。比如针对用户的头像照片，我们首先使用已经用 ImageNet 数据训练好的深度残差网络转化图片得到其语义表示，然后通过聚类算法把图片分为不同的类簇，每个类簇表明了图片所属的类别。通过计算大五人格和类簇的皮尔逊系数，如图 3 展示的与大五人格强烈正相关或者负相关的类簇（每个类簇选取了 2 张图片显示），我们发现了一些有趣的现象：比如外向性得分高的用户喜欢使用包含笑脸的头像，而得分低的用户往往在头像中遮挡了面部表情或者使用侧脸；开放性得分高的用户往往使用和朋友在一起的照片作为头像，而开放性得分低的用户的头像很多是自拍照。我们的实验结果表明，单单使用头像照片，就能使个体性格预测的准确性达到 0.6。我们不仅对每种维度上的行为数据提出了针对性的特征提取策略，而且使用集成学习技术 (ensemble) 有效融合了不同维度的行为数据来提升大五人格预测的准确率，使得个体大五人格预测的准确性达到 0.75 以上。

目前大多数用人工智能方法进行人格预测时，仍然需要采取调查问卷获取的人格数据作为标注，建立模型，虽然可以解决规模性和实时性等问题，但是仍然不能克服调查问卷结果中可能存在的主观性问题。能否在不依赖调查问卷结果的基础上，直接从客观的行为数据中发现人格特质呢？我们在消

⁴ 参见 https://www.microsoft.com/en-us/research/wp-content/uploads/2017/01/WSDM_personality.pdf。

费冲动性和猎奇性这两个细粒度的人格特质上进行了初步研究^[16, 17]。以消费冲动性为例，它在心理学中被定义为用户的购买意愿受到外界刺激影响的程度。基于此原理，我们首先计算用户在每次购买商品时，他们心目中的候选商品对用户产生的刺激影响。比如，我们使用用户浏览的新浪微博作为刺激信息源。近段时间，我们看到大家讨论小米手机微博的次数比较多，那么小米手机这个商品对用户形成的刺激就会比较大。然后，我们构建了一个序列图模型来对用户的购买行为进行建模。这个模型中有一个隐变量，刻画的是用户购买强烈刺激商品的倾向性。通过观察到的行为数据就能从图模型中进行推理，得到这个隐变量的值，而这个隐变量就是我们要从行为数据中发现的消费冲动性。我们比较了行为数据推理得出的消费冲动性和用户调查问卷的结果，发现皮尔逊相关系数达到0.25。如果把用户行为数据的粒度从具体商品提升到商品的类别，皮尔逊相关系数能够达到0.55。这说明，仅仅挖掘用户的行为数据，不依赖调查问卷的标注数据，我们也能在一定程度上获取用户的人格特质。此外，我们还把行为数据推理得到的消费冲动性添加到商品推荐模型中，发现能够有效地提升推荐效果。这为人工智能和人格特质的结合提供了另外一种思路：我们不仅能从行为中预测人格特质，也能利用人格特质预测行为或者影响决策。

小结和展望

几十年来，人格理论和测量方法的不断发展，使得我们能够不断地深入了解不同个体之间的稳定性差异。一方面，在大数据和人工智能时代，被动收集的大量用户行为数据为人格测量提供了新鲜的血液，使得我们不再局限于通过访谈或者调查问卷方法；另一方面，基于行为数据的人工智能方法在人格预测领域仍然属于起步阶段，虽然相关研究已经证实了人工智能方法有效而且切实可行，但是目前的研究基础理论不完整，技术方案尚未形成体系，人格信息的应用不够广泛，这都是该领域目前不成

熟的现状，也同时意味着该领域的研究蕴含了极大的发展空间。人工智能和人格测量的结合可以从以下方面进行更加深入的研究：(1) 需要克服调查问卷的依赖性，直接根据用户行为对用户人格进行预测是该领域需要突破的关键性研究问题；(2) 大数据时代的用户行为可能来自不同平台，是否能够采用相同的建模方法？预测的人格特质结果是否一致？以及如何有效地融合跨领域的异质行为数据并进行用户特征提取，都是亟待解决的问题；(3) 人格预测结果如何帮助产品进行设计？

作为一个交叉学科，这个领域的成功需要计算机科学家、心理学家和社会学家们一起努力。 ■



张富峥

CCF专业会员。微软亚洲研究院副研究员。主要研究方向为行为数据挖掘、推荐系统、深度学习和情感计算。
fuzzhang@microsoft.com



谢 幸

CCF杰出会员、杰出演讲者。微软亚洲研究院高级研究员。主要研究方向为数据挖掘、社会计算和普适计算等。
xingx@microsoft.com

参考文献

- [1] Stagner R. Trait-Names: A Psycho-lexical Study[J]. *Psychological Monographs*, 1936, 47(1):1-171.
- [2] Allport G W. Personality: A psychological interpretation. 1937[M]. New York Henry Holt. 2013.
- [3] Harman H H. *Modern factor analysis*[M]. 1960.
- [4] Cattell R B, Eber H W, Tatsuoka M M. *Handbook for the sixteen personality factor questionnaire (16 PF)*[M]. Institute for Personality and Ability Testing, 1970.
- [5] Fiske D W. Consistency of the factorial structures of personality ratings from different sources[J]. *J. Abnorm. Soc. Psychol.*, 1949, 44(3):329.
- [6] Goldberg L R. An alternative ‘description of personality’: the big-five factor structure[J]. *J. Pers. Soc. Psychol.*, 1990, 59(6):1216.
- [7] Costa P T and McCrae R R. *The NEO personality inventory*[M]. 1985.

- [8] Costa P T, MacCrae R R. *Revised NEO personality inventory (NEO PI-R) and NEO five-factor inventory (NEO-FFI): Professional manual*[M]. Psychological Assessment Resources, Incorporated, 1992.
- [9] Goldberg L. R. The development of markers for the Big-Five factor structure[J]. *Psychol. Assess.* 1992, 4(1):26.
- [10] Howard P. J, Medina P. L and Howard J. M. The big-five locator: A quick assessment tool for consultants and trainers[M]. *Annu. DIEGO-PFEIFFER COMPANY-*, 1996,1:107-122.
- [11] Paunonen S. V, Jackson D. N and Ashton M. C. NPQ Manual: Nonverbal Personality Questionnaire (NPQ) and Five-Factor Nonverbal Personality Questionnaire (FF-NPQ)[M]. Sigma Assessment Systems, 2004.
- [12] 陈基越, 徐建平, 黎红艳等. 五因素取向人格测验的发展与比较[J]. 心理科学进展, 2015, 3: 11.
- [13] Youyou W, Kosinski M, and Stillwell D. Computer-based personality judgments are more accurate than those

- made by humans[J], *Proc. Natl. Acad. Sci.*, 2015, 112(4): 1036-1040.
- [14] Bai S, Gao R, and Zhu T. Determining personality traits from renren status usage behavior, in *Computational Visual Media*, Springer[M], 2012: 226-233.
- [15] Liu X and Zhu T. Deep learning for constructing microblog behavior representation to identify social media user's personality[J]. *PeerJ Comput. Sci.*, 2016, 2:81.
- [16] Zhang F, Yuan N. J, Lian D, and et al. Mining novelty-seeking trait across heterogeneous domains[C]// *Proceedings of the 23rd international conference on World wide web*, 2014:373-384.
- [17] Zhang F, Yuan N. J, Zheng K, and et al. Mining consumer impulsivity from offline and online behavior[C]// *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2015:1281-1292.