



# (12)发明专利

(10)授权公告号 CN 103905486 B

(45)授权公告日 2018.05.15

(21)申请号 201210576344.8

(22)申请日 2012.12.26

(65)同一申请的已公布的文献号  
申请公布号 CN 103905486 A

(43)申请公布日 2014.07.02

(73)专利权人 中国科学院心理研究所  
地址 100101 北京市朝阳区大屯路甲4号

(72)发明人 朱廷劭 张帆 李昂 李一琳

(74)专利代理机构 北京瑞恒信达知识产权代理  
事务所(普通合伙) 11382  
代理人 苗青盛 黄庆芳

(51)Int.Cl.  
H04L 29/08(2006.01)  
G16H 50/30(2018.01)

(56)对比文件

CN 101355504 A,2009.01.28,  
US 2009313041 A1,2009.12.17,  
CN 101355504 A,2009.01.28,  
赵云龙 等.社交网络用户的人格预测与关系强度研究.《第七届(2012)中国管理学年会商务智能分会场论文集(选编)》.2012,  
赵云龙 等.社交网络用户的人格预测与关系强度研究.《第七届(2012)中国管理学年会商务智能分会场论文集(选编)》.2012,  
朱廷劭 等.网络社会中个体人格特征及其行为关系.《兰州大学学报(社会科学版)》.2011,

审查员 陈相玫

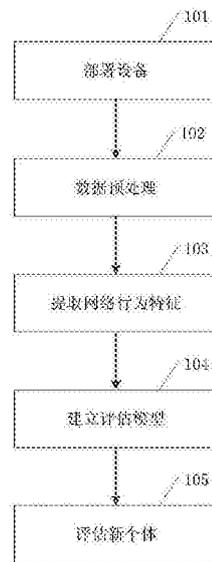
权利要求书2页 说明书9页 附图2页

## (54)发明名称

一种心理健康状态评估方法

## (57)摘要

本申请提供了一种心理健康状态评估方法,包括:1)利用机器学习的方法,基于已知样本中个体网络行为特征和人口统计学特征,建立和训练基于网络行为特征的心理健康状态评估模型;2)获取新个体的网络行为特征和人口统计学特征,根据所述的基于网络行为特征的心理健康状态评估模型,得到该新个体的心理健康状况。本发明消除了个体主观因素对数据收集的影响,有助于更加准确地进行心理健康状态评估。另外,本发明能够更方便地收集心理健康状态评估所需的个体数据,有助于在更广的范围上进行大规模的个体数据采集,实现大规模的心理健康评估。



1. 一种心理状态评估方法,包括下列步骤:

1)、利用机器学习的方法,基于已知样本中个体网络行为特征和人口统计学特征,建立和训练基于网络行为特征的心理状态评估模型;

2)、获取新个体的网络行为特征和人口统计学特征,根据所述的基于网络行为特征的心理状态评估模型,得到该新个体的心理状况;

其中:

所述方法不用于疾病的诊断;

所述网络行为特征是反映个体所使用的网络媒介/服务工具的功能结果和使用路径的特征集合;所述网络行为特征从记录个体的网络日志中提取;以及

所述网络行为特征包括个体的网络信息和时间序列数据,所述个体的网络信息包括:时间信息、各类即时通讯工具信息、邮件信息、所访问网页类别的信息和搜索信息;所述时间序列数据包括:每天的上网时间信息、每天的网络请求个数信息和每天的网页信息;所述时间信息包括:工作日平均每日上网时长和周末平均每日上网时长;所述邮件信息包括是否用客户端收发邮件。

2. 根据权利要求1所述的心理状态评估方法,其特征在于,提取所述网络行为特征的过程包括下列步骤:

11) 利用网络流监控设备,获取个体的网络日志;

12) 对网络日志进行数据预处理;

13) 从网络日志中提取出个体的网络访问记录,所述个体的网络访问记录是以个体为单位存储的结构化文档。

3. 根据权利要求2所述的心理状态评估方法,其特征在于,所述步骤11)中,网络流监控设备以镜像、桥接或者网关方式驳接个体接入互联网的关口控制设备;网络流监控设备根据网络请求的协议类型将获取到的数据包分发到不同的数据表中。

4. 根据权利要求3所述的心理状态评估方法,其特征在于,所述数据表包括:个体身份标识、网络请求的访问时间、客户端IP和端口、服务器IP和端口、传输协议和传输内容。

5. 根据权利要求3所述的心理状态评估方法,其特征在于,所述步骤12)包括下列步骤:

121)、进行数据集成,将来源于针对同一个体的所有网络流监控设备的数据集合起来并统一存储;

122)、进行数据清洗,清除日志中的异常数据,纠正错误,去除冗余数据,标准化日志格式;

123)、进行数据变换,将日志转换成适用于数据挖掘的形式;

所述步骤122)中,对无标题页面做清洗;

所述步骤123)中,清洗后的数据以个体为单位进行聚集,并按时间排序,最后以个体为单位各自单独保存为网络访问记录;所述网络访问记录采用XML标准作为文档的存储格式。

6. 根据权利要求1所述的心理状态评估方法,其特征在于,所述步骤1)、2)中,所述基于网络行为特征的心理状态评估模型采用的数学模型是回归模型、基于分类器的模型或者M5P回归模型。

7. 根据权利要求1所述的心理状态评估方法,其特征在于,所述所访问网页类别的信息包括:访问数量排名前列的网页类别、经常访问的网页类别个数、用过的BBS站点个数、用过

的社交网络个数和用过的微博个数。

8. 根据权利要求1所述的心理状态评估方法,其特征在於,所述搜索信息包括:搜索过的类别个数、经常搜索的类别个数、用过的搜索引擎个数和常用的搜索引擎的个数;

所述每天的上网时间信息包括:总时长、白天的上网时长、晚上的上网时长、上网时间段个数、上线时刻、下线时刻、活跃时刻、是否半夜上网和是否深夜上网。

9. 根据权利要求1所述的心理状态评估方法,其特征在於,所述每天的网络请求个数信息包括:总请求数、每个类型的请求数、即时通讯的聊天对象个数和处理的信息数、对网页分类后的各类网页请求个数、各类网页中的第一个网页;

所述每天的网页信息包括:所访问的网站个数、网页类别数、直接或由收藏夹访问的网页数、搜索内容的类别数、使用的搜索引擎数和搜索关键字平均长度;

所述网络行为特征还包括上网天数的统计数据,所述上网天数的统计数据包括:总天数、上网天数、不上网天数和深夜上网天数。

10. 一种建立和训练基于网络行为特征的心理状态评估模型的方法,

利用机器学习的方法,基于已知样本中个体网络行为特征和人口统计学特征,建立和训练基于网络行为特征的心理状态评估模型;所述网络行为特征是反映个体所使用的网络媒介/服务工具的功能结果和使用路径的特征集合;其中,所述方法不用于疾病的诊断;所述网络行为特征包括个体的网络信息和时间序列数据,所述个体的网络信息包括:时间信息、各类即时通讯工具信息、邮件信息、所访问网页类别的信息和搜索信息;所述时间序列数据包括:每天的上网时间信息、每天的网络请求个数信息和每天的网页信息;所述时间信息包括:工作日平均每日上网时长和周末平均每日上网时长;

所述邮件信息包括是否用客户端收发邮件。

## 一种心理健康状态评估方法

### 技术领域

[0001] 本发明涉及心理健康状态评估和互联网技术领域,具体地说,本发明涉及一种心理健康状态评估方法。

### 背景技术

[0002] 当前,在世界范围内,心理健康问题已经成为了导致个体“失能”(disability)的首要原因,其消极性后果占有所有疾病性危害的37%,结合其“全球性”(global)、“长期性”(chronic)和“流行性”(prevalent)的发展趋势来看,心理健康问题的存在不仅会造成社会资源的巨大消耗与浪费,加剧了整个社会的负担,而且也直接威胁到个体的日常生活质量和主观幸福感。具体到我国的实际情况,根据调查研究结果显示,民众心理健康的整体水平同样不容乐观,心理健康问题的整体发生率达到17.5%。由于我国拥有庞大的人口基数,所以采用高效的应对措施来解决心理健康问题的需求更为迫切。

[0003] 通常,提供心理健康服务的第一步就是对个体进行“心理健康状态评估”。心理健康状态评估涉及到针对个体的特异性行为症状进行收集、分析、综合、判断,具体可采用“访谈”、“问卷测验”、“投射测验”等多种技术手段实现(可参考文献:Anatasi A, U.S..Psychological testing.New Jersey:Prentice Hall,1997.)。其中,近些年来“问卷测验”因其在工具编制、施测操作、结果解释等方面所具有的结构性和客观性获得了相对更广泛的临床应用。“问卷测验”技术通常会借助于一些综合性的心理健康测评工具(比如MMPI-2、SCL-90问卷)从多个角度来综合考量个体的心理健康状态。

[0004] 心理测量学(Psychometrics)上,“问卷测验”所用到的评定量表(rating scales)是用来量化观察中所得印象的一种测量工具,它是心理状态评估中收集资料的重要手段之一(可参考文献:汪向东,王希林,马弘.心理卫生量表评定手册.中国心理卫生杂志出版社,1999.)。评定量表由若干项目组成,每一个项目都可以看作是对一系列行为特征的抽象化描述,而这些行为特征与心理特征存在着一定的关系。用量表评估心理健康状态时,首先,需要个体根据自身情况与项目内容的契合度填写量表;其次,评定员根据量表手册提供的评分方法计算得出评分;最后,评定员根据测评结果,提出结论,并对其意义进行解释,以文字或口头形式报告给个体。

[0005] “问卷测验”技术已被广泛应用于心理健康状态评估,然而,其自身仍然存在一些不足之处,有待改进。这些不足之处包括:

[0006] (1)收集到的数据准确度会受到被试主观因素的影响。因为“问卷测验”技术使个体在研究中同时担任着“被观察客体”和“观察主体”两种角色,当个体在填写量表时,其答案不可避免地会受到社会赞许心理、个体认知能力的影响。虽然通过一些控制手段(比如,效度量表、实验设计)可以排查其中某些主观偏差因素,但是在同类研究中反复实施类似的控制却十分繁琐低效,且无法排除部分无效数据(比如,错答、漏答、盲答)。

[0007] (2)收集到的数据细节和规模有限。首先,由于受时空和一些社会条件的限制,每次实验很难召集到数量充足的个体填写量表,而且对同一批个体进行跟踪测量也比较困

难。其次,为了方便个体完成调查,量表的题目不可能设计的非常多,从而收集不到个体行为样本的更多细节。此外,量表的发放、收集与录入也给大规模采集数据带来了不利的影响。

## 发明内容

[0008] 为克服现有的缺陷,本发明提出一种能够消除个体主观因素影响,且便于大规模采集数据的心理健康状态评估方法。

[0009] 根据本发明的一个方面,提出了一种心理健康状态评估方法,包括下列步骤:

[0010] 1) 利用机器学习的方法,基于已知样本中个体网络行为特征和人口统计学特征,建立和训练基于网络行为特征的心理健康状态评估模型;

[0011] 2) 获取新个体的网络行为特征和人口统计学特征,根据所述的基于网络行为特征的心理健康状态评估模型,得到该新个体的心理健康状况。

[0012] 其中,所述网络行为特征是反映个体所使用的网络媒介/服务工具的功能结果和使用路径的特征集合。

[0013] 其中,所述步骤1)、2)中,所述网络行为特征从记录个体的网络日志中提取。

[0014] 其中,提取所述网络行为特征的过程包括下列步骤:

[0015] 11) 利用网络流监控设备,获取个体的网络日志;

[0016] 12) 对网络日志进行数据预处理;

[0017] 13) 从网络日志中提取出个体的网络访问记录,所述个体的网络访问记录是以个体为单位存储的结构化文档。

[0018] 其中,所述步骤11)中,网络流监控设备以镜像、桥接或者网关方式驳接个体接入互联网的关口控制设备。

[0019] 其中,网络流监控设备根据网络请求的协议类型将获取到的数据包分发到不同的数据表中。

[0020] 其中,所述数据表包括:个体身份标识、网络请求的访问时间、客户端IP和端口、服务器IP和端口、传输协议和传输内容。

[0021] 其中,所述步骤12)包括下列子步骤:

[0022] 121) 进行数据集成,将来源于针对同一个体的所有网络流监控设备的数据集合起来并统一存储;

[0023] 122) 进行数据清洗,清除日志中的异常数据,纠正错误,去除冗余数据,标准化日志格式;

[0024] 123) 进行数据变换,将日志转换成适用于数据挖掘的形式。

[0025] 其中,所述步骤122)中,对无标题页面做清洗。

[0026] 其中,所述步骤123)中,清洗后的数据以个体为单位进行聚集,并按时间排序,最后以个体为单位各自单独保存为网络访问记录。

[0027] 其中,所述步骤123)中,所述网络访问记录采用XML标准作为文档的存储格式。

[0028] 其中,所述步骤1)、2)中,所述基于网络行为特征的心理健康状态评估模型采用的数学模型是回归模型或者基于分类器的模型。

[0029] 其中,所述步骤1)、2)中,所述基于网络行为特征的心理健康状态评估模型采用的

数学模型是M5P回归模型。

[0030] 其中,所述网络行为特征包括个体的网络信息和时间序列数据,所述个体的网络信息包括:时间信息、各类即时通讯工具信息、邮件信息、所访问网页类别的信息和搜索信息;所述时间序列数据包括:每天的上网时间信息、每天的网络请求个数信息和每天的网页信息。

[0031] 其中,所述时间信息包括:工作日平均每日上网时长和周末平均每日上网时长。

[0032] 其中,所述各类即时通讯工具信息包括:分别拥有的账号个数、各类即时通讯工具的好友个数和各类即时通讯工具经常联系的好友个数。

[0033] 其中,所述邮件信息包括是否用客户端收发邮件。

[0034] 其中,所述所访问网页类别的信息包括:访问数量排名前列的网页类别、经常访问的网页类别个数、用过的BBS站点个数、用过的社交网络个数和用过的微博个数。

[0035] 其中,所述搜索信息包括:搜索过的类别个数、经常搜索的类别个数、用过的搜索引擎个数和常用的搜索引擎的个数。

[0036] 其中,所述每天的上网时间信息包括:总时长、白天的上网时长、晚上的上网时长、上网时间段个数、上线时刻、下线时刻、活跃时刻、是否半夜上网和是否深夜上网。

[0037] 其中,所述每天的网络请求个数信息包括:总请求数、每个类型的请求数、即时通讯的聊天对象个数和处理的信息数、对网页分类后的各类网页请求个数、各类网页中的第一个网页。

[0038] 其中,所述每天的网页信息包括:所访问的网站个数、网页类别数、直接或由收藏夹访问的网页数、搜索内容的类别数、使用的搜索引擎数和搜索关键字平均长度。

[0039] 其中,所述网络行为特征还包括上网天数的统计数据,所述上网天数的统计数据包括:总天数、上网天数、不上网天数和深夜上网天数。

[0040] 与现有技术相比,本发明具有下列技术效果:

[0041] 1、本发明消除了个体主观因素对数据收集的影响,有助于更加准确地进行心理健康状态评估。

[0042] 2、本发明能够更方便地收集心理健康状态评估所需的个体数据,有助于在更广的范围上进行大规模的个体数据采集,实现大规模的心理健康评估。

## 附图说明

[0043] 图1是本发明一个实施例的心理健康状态评估方法的流程示意图;

[0044] 图2是本发明一个实施例中描述用户网络行为过程的SMEE模型示意图;

[0045] 图3是本发明一个实施例中各类网络行为特征、人口统计学特征、心里特征的关系示意图。

## 具体实施方式

[0046] 为了使发明的目的、技术方案和优点更加清楚明白,下面举具体实施例,对本发明作进一步详细的说明。

[0047] 为便于理解,首先简要介绍本发明所依据的科学原理。

[0048] 在心理测量学中,评估心理健康状态都通过个体的外显的“行为样本”(behavior

sample) 指标进行间接测量。因为个体的行为是受其心理状态支配和影响的,心理健康状况的差异可以从行为的差异上体现出来,所以可以借助能从部观测到的“行为样本”来获知个体的心理健康状况。

[0049] 而另一方面,网络社会已经发展成熟。据CNNIC《第28次中国互联网发展状况统计报告》显示,截至2011年6月底,我国互联网普及率为36.2%,网民总数达到4.85亿,其中学生群体所占比例高达29.9%。网络行为已经成为人们行为主体一个不可或缺的重要组成部分。随着互联网的不断普及,网络社会的日趋完善,人们的学习、工作、生活已紧密的和互联网结合起来,人们使用互联网的时间越来越多,网络生活已经成为人们日常生活的一部分。同时,网络行为也成为人们行为主体的一个重要组成部分。因此,在网络社会高度发展的今天,使用网络行为评估心理健康状态已有其实现基础。

[0050] 基于上述原理,根据本发明的一个实施例,提供了一种基于网络行为的心理健康状态评估方法,它首先利用现有的机器学习的方法,基于已知的样本来寻找网络行为特征与心理健康状况之间的联系,从而建立和训练基于网络行为特征的心理健康状态评估模型,然后再获取新个体的网络行为特征,根据训练得到的模型和新个体的网络行为特征,可以得到新个体的心理健康状况。

[0051] 图1为一个实施例中基于网络行为特征的心理健康状态评估的流程示意图,具体步骤如下:

[0052] 步骤101,部署网络流监控设备,开启记录网络日志记录功能,获取记录个体网络行为的网络日志。

[0053] 在一个实施例中,将网络流监控设备以镜像的方式和网关驳接,所有流经网关的数据都会被复制一份副本发送至网络流监控设备,然后网络流监控设备分析获得的数据包,将其转换为网络请求后存储为结构化的日志文档,并存储到数据库中。除了镜像的方式外,在另一些实施例中,网络流监控设备的部署方式还可以是桥接、网关等多种常规驳接方法,这是本领域普通技术人员易于理解的。

[0054] 网络流监控设备能获取流经网关的数据包。而网关为个体接入互联网的关口控制设备,个体产生的所有网络请求必须经过网关设备才能接入互联网。在一个实施例中,网络流监控设备能分析获取到的数据包,并将其转换为网络请求后以日志形式记录下来,所述的日志为记录有每一条网络请求的结构化文档。该实施例中,网络流监控设备根据网络请求的协议类型分发到不同的数据表中,不同协议的数据表结构并不完全相同,但都包含有网络请求的一些主要内容,如:个体身份标识、网络请求的访问时间、客户端IP和端口、服务器IP和端口、传输协议、传输内容。所有的数据表按天对表进行切分,表名为“协议名\_日期”。例如,2012年1月1日中所有协议类型为http的网络请求将保存至表名为http\_20120101的数据表中,其字段结构如表1所示。

[0055] 表1

[0056]

id	数据表的逻辑编号	request_method	请求方式, 如 GET/POST 等
client_identity	个体身份标识	refer	前向页面的 url
request_time	网络请求产生时间	title	页面标题
client_ip	客户端 ip	url	页面的 url
upload_size	上传量	transfer_encoding	传输编码, 如 chunked
download_size	下载量	content_coding	内容编码, 如 gzip

[0057] 步骤102, 从网络流监控设备获取日志, 并进行数据预处理, 从中提取出个体的网络访问记录。

[0058] 一个实施例中, 预处理包括: 数据集成, 其目的是将来源于多个网络流监控设备的数据集合起来并统一存储, 即建立初始的数据仓库; 数据清洗, 其目的是清除日志中的异常数据, 纠正错误, 去除冗余数据, 标准化日志格式; 以及数据变换, 其目的是通过平滑聚集, 数据概化, 规范化等方式将日志转换成适用于数据挖掘的形式。其中, 当个体的网络行为只有一个数据来源, 即每个个体只通过一个网络流监控设备上网。在对数据集进行预处理时可以跳过数据集成步骤, 直接进入数据清理步骤, 然后再执行数据变换步骤。

[0059] 数据清理的规则需要根据实际业务需求进行调整, 在本实施例中, 对无标题页面做清洗。无标题页面绝大多数是对个体心理健康状态评估无实质价值的页面, 例如: 页面本身产生的数据请求, 应用程序的网络数据交换请求等), 网关登陆离线记录, 广告页面, 站点数据统计请求, 错误请求 (404等), 冗余请求 (例如 instant 搜索技术产生的请求, 因此清洗无标题页面能够去除冗余数据, 有助于更准确地反应个体的真实网络行为。

[0060] 数据变换负责将清洗后的数据以个体为单位进行聚集, 并按时间升序规则进行排序, 最后以个体为单位各自单独保存为网络访问记录。为使所保存的网络访问记录便于交换和扩展, 本实施例中采用 XML 标准作为文档的存储格式。

[0061] 步骤103, 结合网络行为指标体系, 提取个体的网络行为特征。

[0062] 网络行为指标体系是从个体具体的网络访问记录中抽象出的更一般和更具代表性的网络行为特征体系。

[0063] 一个实施例中, 所述的网络行为指标体系根据描述用户网络行为过程的 SMEE 模型 (见图2) 得出。其中, 利用网络媒介/服务工具来达成自己的目标和期望是一个决策过程。决策过程指的是选择行动方案的全过程, 即在指定环境中有效选择可用于达成目标的方案。在网络使用过程中, 用户试图找到最符合自身实际心理需求的关于媒介/服务工具的选择/应用模式, 这表现为在应用各种媒介/服务工具时找到其各自对应的使用效应, 通过结果 (是否满足了心理需求) 反馈, 更多更快地使其在实际应用时产生的意外效应转化为预期效应, 从而最终选择那个最可能代表了理想预期效应的网络媒介/服务工具。供选择的网络使用方案具体包括了网络媒介/服务工具的不同使用路径与功能结果的组合, 而行为主体的个体特征则为多种组合的网络使用方案的选择性执行提供了依据 (赋予权重和编排执行顺序), 因此网络行为可以被看做是针对多种网络使用方案的选择性执行结果或执行的外显, 它是个体特征 (人口统计学特征、心理特征)、网络媒介/服务工具功能结果、网络媒介/服务

工具使用路径的函数。其中,人口统计学特征、网络媒介/服务工具的功能结果和使用路径均是外显可观察的,因此可以借助用户针对行为方案的具体选择结果(选择的媒介/工具的功能结果和使用路径),结合其自身的人口统计学资料,来还原用户的心理特征(见图3)。人口统计学特征、网络媒介/服务工具的功能结果特征和使用路径特征可以作为组织网络行为指标体系的框架。

[0064] 在一个优选实施例中,从访问时间、访问协议类别、访问内容类别等方面提取了100多个网络行为特征,这些特征可以反映个体所使用的网络媒介/服务工具的功能结果和使用路径,这些特征再与个体的人口统计学信息结合在一起作为该个体的行为特征集合。特征集合的详细内容如表2所示。

[0065] 表2

[0066]	个人信息	人口统计学信息		性别	
				年龄	
				受教育程度	
				专业	
				民族	
				独生子女	
				家乡	
				家庭月收入	
	网络信息	上网时间	工作日平均每日上网时长		
			周末平均每日上网时长		
		通讯	即时通讯	qq	拥有的账号个数
				fetion	拥有的账号个数
					好友个数
				经常联系的好友个数	
msn	拥有的账号个数				
	好友个数				

[0067]

					经常联系的好友个数	
			email	是否用客户端收发邮件		
		网页	访问数量排名前 5 的网页类别			
			经常访问的网页类别个数			
			用过的 BBS 站点个数			
			用过的社交网络个数			
			用过的微博个数			
			搜索	搜索过的类别个数		
				经常搜索的类别个数		
				用过的搜索引擎个数		
				常用的搜索引擎的个数		
			其他	是否使用 FTP		
是否使用 TELNET						
具体网络行为特征	时间序列数据（每天）	上网时间	总时长			
			白天的上网时长			
			晚上的上网时长			
			上网时间段个数			
			上线时刻			
			下线时刻			
			活跃时刻			
			是否半夜上网			
			是否深夜上网			
		总请求数	总请求数			
			白天的请求数			
			晚上的请求数			
		每个类型的请求数	网页请求数			
			聊天请求数			
			HTTPS 请求个数			
MAIL 请求个数						
FTP 请求个数						
网络请求个数	fetion	聊天对象个数				
		处理的信息总数				
		接受信息数				
		发送信息数				
	即时通讯	msn	聊天对象个数			
			处理的信息总数			
			接受信息数			
			发送信息数			
	qq	处理的信息总数				
接受信息数						
发送信息数						
网页分类	网康分类	第 1 类				

[0068]				其余	
				搜索请求个数	
				社交网站请求数	
				微博请求个数	
			网页分类详情	搜索内容	第1类
					其余
				搜索引擎	第1个
					其余
				社交网站	第1个
					其余
				微博	第1个
					其余
	网页	概况	网站个数		
			网页类别数		
			直接或由收藏夹访问的网页数		
		搜索	搜索内容的类别数		
使用的搜索引擎数					
搜索关键字平均长度					
统计数据	上网天数	总天数			
		上网天数			
		不上网天数			
		深夜上网天数			

[0069] 步骤104,利用机器学习的方法,建立基于网络行为特征的心理健康状态评估模型。

[0070] 在本实施例中,采用一种回归模型M5P来建立心理健康状态评估模型,M5P是一种结合了线性回归的决策树模型,它可以用来预测连续值变量。在本实施例中,M5P将个体的网络行为特征视为样本的特征向量,将个体的心理健康状态得分视为需要预测的连续值变量,通过相应的机器学习训练算法,得出心理健康状态评估模型,其中训练所需样本,即训练所需的个体的心理健康状态事先通过评定量表测量和收集。在本实施例中,使用机器学习工具包weka来建立基于M5P的心理健康状态预测模型。当然,在其它实施例中,也可以采用基于分类器的机器学习方法。

[0071] 步骤105,将新个体的网络行为特征输入到心理健康状态评估模型,得到其心理健康状态评估结果。

[0072] 其中,新个体的网络行为特征可通过前述步骤101、102、103得出。心理健康状态评估模型是在步骤104中训练得到的数学模型。

[0073] 上述实施例给出了一种基于网络行为的心理健康状态评估方案,从个体的网络访问日志中提取出其行为特征,并用机器学习的方法建立通过网络行为特征评估心理健康状态的模型,以此评估新个体的心理健康状态。从而避免了传统“问卷测验”技术带来的主观因素的影响,更方便、准确、客观地收集心理健康状态评估所需的个体数据,同时,凭借网络的优势,可以在更广的范围上进行大规模的个体数据采集,实现大规模的心理健康评估。

[0074] 以上所述仅为本发明的较佳实施例,并不用以限制本发明,凡在本发明的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

[0075] 最后应说明的是,以上实施例仅用以描述本发明的技术方案而不是对本技术方法进行限制,本发明在应用上可以延伸为其他的修改、变化、应用和实施例,并且因此认为所有这样的修改、变化、应用、实施例都在本发明的精神和教导范围内。

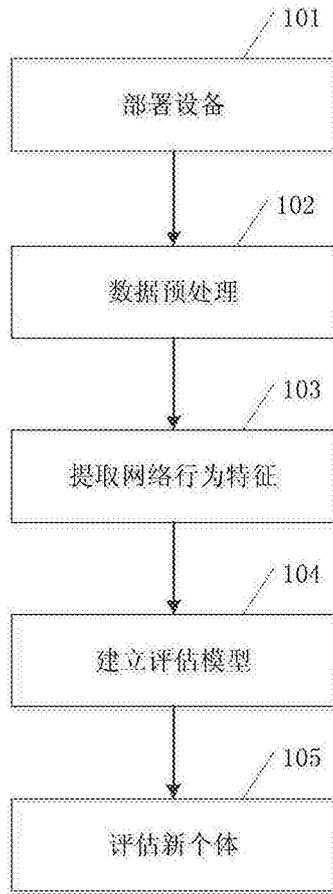


图1

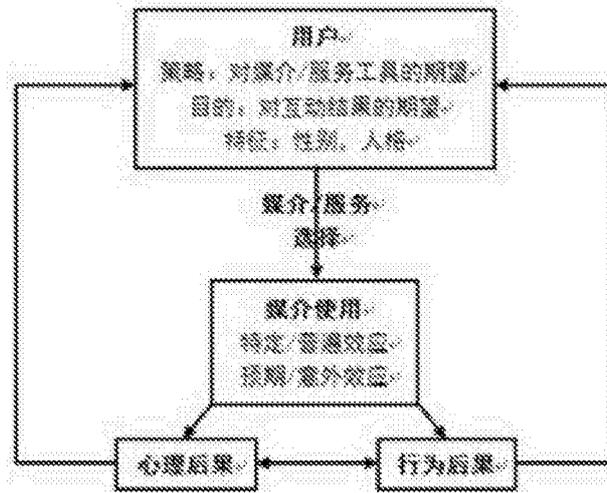


图2

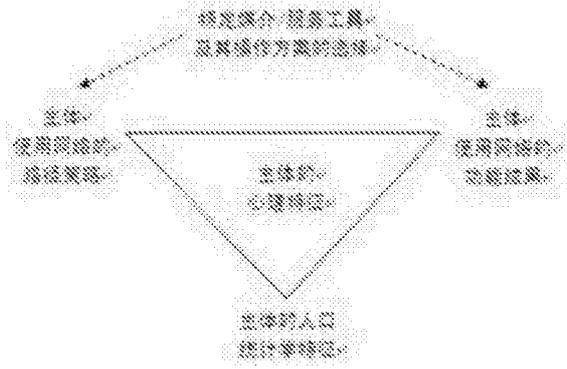


图3