



(12) 发明专利

(10) 授权公告号 CN 107958317 B

(45) 授权公告日 2021.12.14

(21) 申请号 201610901108.7

(22) 申请日 2016.10.17

(65) 同一申请的已公布的文献号
申请公布号 CN 107958317 A

(43) 申请公布日 2018.04.24

(73) 专利权人 腾讯科技(深圳)有限公司
地址 518000 广东省深圳市福田区振兴路
赛格科技园2栋东403室

(72) 发明人 郑博 刘志斌 刘日佳

(74) 专利代理机构 深圳市深佳知识产权代理事
务所(普通合伙) 44285
代理人 王仲凯

(51) Int.Cl.
G06Q 10/06 (2012.01)

(56) 对比文件

CN 105138521 A, 2015.12.09

CN 103064970 A, 2013.04.24

US 2013311220 A1, 2013.11.21

US 2015254596 A1, 2015.09.10

朱小宁.“支持任务推送的众包系统的研究与实现”.《中国优秀硕士学位论文全文数据库》.2015,第I138-357页.

审查员 洪利燕

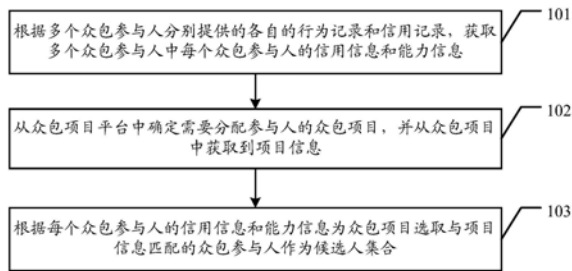
权利要求书4页 说明书18页 附图4页

(54) 发明名称

一种众包项目中选取众包参与人的方法和装置

(57) 摘要

本发明公开一种众包项目中选取众包参与人的方法和装置,用于为众包项目平台中的众包项目选取众包参与人,并能够保证筛选的可靠性。本发明实施例提供的方法中,根据多个众包参与人分别提供的各自的行为记录和信用记录,获取所述多个众包参与人中每个众包参与人的信用信息和能力信息;从众包项目平台中确定需要分配参与人的众包项目,并从所述众包项目中获取到项目信息,所述项目信息包括:所述众包项目的项目任务以及所述项目任务的内容信息;根据所述每个众包参与人的信用信息和能力信息为所述众包项目选取与所述项目信息匹配的众包参与人作为候选人集合,所述候选人集合包括:从所述多个众包参与人中筛选出的用于执行所述项目任务的众包参与人。



1. 一种众包项目中选取众包参与人的方法,其特征在于,包括:

根据多个众包参与人分别提供的各自的行为记录和信用记录,获取所述多个众包参与人中每个众包参与人的信用信息和能力信息;

从众包项目平台中确定需要分配参与人的众包项目,并从所述众包项目中获取到项目信息,所述项目信息包括:所述众包项目的项目任务以及所述项目任务的内容信息;

根据所述每个众包参与人的信用信息和能力信息为所述众包项目选取与所述项目信息匹配的众包参与人作为候选人集合,所述候选人集合包括:从所述多个众包参与人中筛选出的用于执行所述项目任务的众包参与人;

其中,所述从所述众包项目中获取到项目信息,包括:

确定所述众包项目的项目任务,获取所述项目任务的内容信息,所述内容信息包括:所述众包项目的名称、类型、知识领域范围、时间、费用、质量、资源、风险;

根据所述项目任务的内容信息估计所述项目任务所需的成本和所述项目任务的完成要求,其中,所述项目任务所需的成本包括:人力成本、资金成本和时间成本,所述人力成本包括:所述项目任务所需的专业技能类型、人员数量和专业技能水平,所述资金成本包括:所述项目任务的项目报酬、项目沟通渠道所需要的资金、项目资源提供成本、项目开销成本,所述时间成本包括:项目沟通成本和项目工时成本,所述完成要求包括:完成可靠度、完成质量、未完成和延期的损失;

所述根据所述每个众包参与人的信用信息和能力信息为所述众包项目选取与所述项目信息匹配的众包参与人作为候选人集合,包括:

根据所述每个众包参与人的信用信息和能力信息、所述项目任务的内容信息,预测所述每个众包参与人的项目接受可能性指标、项目中途放弃可能性指标;

根据所述每个众包参与人的能力信息和所述每个众包参与人的项目接受可能性指标,从所述多个众包参与人中查找到符合所述项目任务所需的人力成本、且符合所述项目任务的完成质量的第一众包参与人子集;

根据所述每个众包参与人的信用信息从所述多个众包参与人中查找到符合所述项目任务的完成可靠度的第二众包参与人子集;

根据所述每个众包参与人的能力信息和所述每个众包参与人的项目中途放弃可能性指标,从所述多个众包参与人中查找到符合所述项目任务所需的时间成本、资金成本、且符合所述项目任务的未完成和延期的损失的第三众包参与人子集;

根据所述第一众包参与人子集、所述第二众包参与人子集、所述第三众包参与人子集确定出与所述项目信息匹配的众包参与人作为候选人集合;

其中,所述根据多个众包参与人分别提供的各自的行为记录和信用记录,获取所述多个众包参与人中每个众包参与人的信用信息和能力信息,包括:

收集多个众包参与人的各自的用户信息,所述用户信息包括:历史众包行为记录、社交网络行为记录、社会信用记录和个人认证信息,其中,所述历史众包行为记录通过多个众包参与人在所述众包项目平台上的各自账户得到,所述社交网络行为记录通过多个众包参与人各自的社交网络账户得到,所述社会信用记录通过多个众包参与人各自的社会信用的授权查询得到,所述个人认证信息通过多个众包参与人各自提供的个人认证资料得到;

根据所述多个众包参与人的各自的历史众包行为记录、社交网络行为记录和个人认证

信息计算每个众包参与人的个人通用素质能力和专业知识能力；

根据所述多个众包参与人的各自的社会信用记录和个人认证信息计算每个众包参与人的信用水平指标。

2. 根据权利要求1所述的方法,其特征在于,所述根据所述每个众包参与人的信用信息和能力信息为所述众包项目选取与所述项目信息匹配的众包参与者作为候选人集合之后,所述方法还包括:

将选取的所述候选人集合通知给所述众包项目平台,由所述众包项目平台根据所述候选人集合对所述项目任务进行分发;

根据所述候选人集合中的各个众包参与人的工作进度对所述项目任务的项目进度进行评估,根据评估结果预测所述项目任务的项目进度是否符合预期,再根据预测结果调整所述候选人集合。

3. 根据权利要求1所述的方法,其特征在于,所述根据所述多个众包参与人的各自的社会信用记录和个人认证信息计算每个众包参与人的信用水平指标之后,所述方法还包括:

计算所述多个众包参与人的各自的个人认证信息与收集到的所述用户信息之间的一致性指标,并保存得到的一致性计算结果。

4. 根据权利要求3所述的方法,其特征在于,所述根据所述每个众包参与人的信用信息从所述多个众包参与者中查找到符合所述项目任务的完成可靠度的第二众包参与者子集,包括:

若所述一致性计算结果表示众包参与人的个人认证信息与收集到的用户信息存在不一致时,将所述一致性计算结果存储为反欺诈识别信息;

根据所述每个众包参与人的信用信息和所述反欺诈识别信息从所述多个众包参与者中查找到符合所述项目任务的完成可靠度的第二众包参与者子集。

5. 一种众包项目中选取众包参与人的装置,其特征在于,包括:

参与者信息处理模块,用于根据多个众包参与者分别提供的各自的行为记录和信用记录,获取所述多个众包参与者中每个众包参与人的信用信息和能力信息;

众包项目信息处理模块,用于从众包项目平台中确定需要分配参与人的众包项目,并从所述众包项目中获取到项目信息,所述项目信息包括:所述众包项目的项目任务以及所述项目任务的内容信息;

众包参与者选取模块,用于根据所述每个众包参与人的信用信息和能力信息为所述众包项目选取与所述项目信息匹配的众包参与者作为候选人集合,所述候选人集合包括:从所述多个众包参与者中筛选出的用于执行所述项目任务的众包参与者;

其中,所述众包项目信息处理模块,包括:

项目任务处理模块,用于确定所述众包项目的项目任务,获取所述项目任务的内容信息,所述内容信息包括:所述众包项目的名称、类型、知识领域范围、时间、费用、质量、资源、风险;

项目信息分析模块,用于根据所述项目任务的内容信息估计所述项目任务所需的成本和所述项目任务的完成要求,其中,所述项目任务所需的成本包括:人力成本、资金成本和时间成本,所述人力成本包括:所述项目任务所需的专业技能类型、人员数量和专业技能水平,所述资金成本包括:所述项目任务的项目报酬、项目沟通渠道所需要的资金、项目资源

提供成本、项目开销成本,所述时间成本包括:项目沟通成本和项目工时成本,所述完成要求包括:完成可靠度、完成质量、未完成和延期的损失;

所述众包参与人选取模块,包括:

用户意愿分析模块,用于根据所述每个众包参与人的信用信息和能力信息、所述项目任务的项目信息,预测所述每个众包参与人的项目接受可能性指标、项目中途放弃可能性指标;

人力需求与参与人能力匹配模块,用于根据所述每个众包参与人的能力信息和所述每个众包参与人的项目接受可能性指标,从所述多个众包参与人中查找到符合所述项目任务所需的人力成本、且符合所述项目任务的完成质量的第一众包参与人子集;

完成可靠度与参与人信用匹配模块,用于根据所述每个众包参与人的信用信息从所述多个众包参与人中查找到符合所述项目任务的完成可靠度的第二众包参与人子集;

项目进度保证模块,用于根据所述每个众包参与人的能力信息和所述每个众包参与人的项目中途放弃可能性指标,从所述多个众包参与人中查找到符合所述项目任务所需的时间成本、资金成本、且符合所述项目任务的未完成和延期的损失的第三众包参与人子集;

候选人集合确定模块,用于根据所述第一众包参与人子集、所述第二众包参与人子集、所述第三众包参与人子集确定出与所述项目信息匹配的众包参与人作为候选人集合;

其中,所述参与人信息处理模块,包括:

收集模块,用于收集多个众包参与人的各自的用户信息,所述用户信息包括:历史众包行为记录、社交网络行为记录、社会信用记录和个人认证信息,其中,所述历史众包行为记录通过多个众包参与人在所述众包项目平台上的各自账户得到,所述社交网络行为记录通过多个众包参与人各自的社交网络账户得到,所述社会信用记录通过多个众包参与人各自的社会信用的授权查询得到,所述个人认证信息通过多个众包参与人各自提供的个人认证资料得到;

能力计算模块,用于根据所述多个众包参与人的各自的历史众包行为记录、社交网络行为记录和个人认证信息计算每个众包参与人的个人通用素质能力和专业知识能力;

信用计算模块,用于根据所述多个众包参与人的各自的社会信用记录和个人认证信息计算每个众包参与人的信用水平指标。

6. 根据权利要求5所述的装置,其特征在于,所述众包项目中选取众包参与人的装置还包括:

众包项目跟踪模块,用于所述众包参与人选取模块根据所述每个众包参与人的信用信息和能力信息为所述众包项目选取与所述项目信息匹配的众包参与人作为候选人集合之后,将选取的所述候选人集合通知给所述众包项目平台,由所述众包项目平台根据所述候选人集合对所述项目任务进行分发;根据所述候选人集合中的各个众包参与人的工作进度对所述项目任务的项目进度进行评估,根据评估结果预测所述项目任务的项目进度是否符合预期,再根据预测结果调整所述候选人集合。

7. 根据权利要求5所述的装置,其特征在于,所述信用计算模块,还用于根据所述多个众包参与人的各自的社会信用记录和个人认证信息计算每个众包参与人的信用水平指标之后,计算所述多个众包参与人的各自的个人认证信息与收集到的所述用户信息之间的一致性指标,并保存得到的一致性计算结果。

8. 根据权利要求7所述的装置,其特征在於,所述完成可靠度与参与人信用匹配模块,具体用于若所述一致性计算结果表示众包参与人的个人认证信息与收集到的用户信息存在不一致时,将所述一致性计算结果存储为反欺诈识别信息;根据所述每个众包参与人的信用信息和所述反欺诈识别信息从所述多个众包参与人中查找到符合所述项目任务的完成可靠度的第二众包参与人子集。

一种众包项目中选取众包参与人的方法和装置

技术领域

[0001] 本发明涉及计算机技术领域,尤其涉及一种众包项目中选取众包参与人的方法和装置。

背景技术

[0002] 随着互联网的兴起,互联网上充斥着用户产生的内容。在较早的时期,这些内容往往是无组织的个人内容,但随着网络社群的发展,越来越多的自组织内容产生行为开始出现,例如论坛和问答社区的精华板块,以及网络游戏中的活动等等。这些自组织内容虽然能够满足其他人通过网络获取知识或娱乐的需求,但尚不足以构成生产力。为了能够有效组织网络社群当中的内容生产和信息获取能力,众包作为一种有效的组织方式被提出,并且在实践中诞生了Wikipedia和Github等成功的协作平台,产生了能够影响世界的项目和成果,展示了众包的优势和能力。

[0003] 常见的众包平台通常需要一个或多个项目提出者,例如Wikipedia的词条提出人,然后有众多的参与者加入完成这一项目,例如Wikipedia用户对于词条的编辑和修订。为了能够有效控制不合理的行为,例如错误的知识或者营销信息,还会引入诸如质疑、不实信息挑战等机制进行修正和控制,如果符合条件的参与人的数量能够达到理想水平,且目的和动机与项目一致,那么众包是一种很有效的信息产生、收集、处理、总结的机制。

[0004] 但众包对于某些难度较大的工作通常难以满足这一条件,例如论文的评审过程,设计方案的征集过程等,常常面临对于参与人员专业技能要求过高的问题,从而难以在短时间内召集到足够多的参与人。另外一个问题则是网络社群的匿名性,这使得对于错误产出和恶意破坏的惩罚难以实行,使得参与人员的动机难以保证,从而影响众包项目的完成质量和进度,甚至造成项目的失败。

[0005] 在现有技术中对众包参与人进行选取的过程中,搭建好众包项目平台之后,通常记录众包项目提供方和参与方的信息,为双方提供中介和担保服务,通过平台人工查询匹配的方式为众包项目找到能够满足专业要求的参与人。这种众包参与人的选取方案中,项目平台往往需要收取众包项目提供方和参与方一定的保证金来提供担保,提高了众包参与的门槛,但也无法对众包参与人进行有效筛选。

发明内容

[0006] 本发明实施例提供了一种众包项目中选取众包参与人的方法和装置,用于为众包项目平台中的众包项目选取众包参与人,并能够保证筛选的可靠性。

[0007] 为解决上述技术问题,本发明实施例提供以下技术方案:

[0008] 第一方面,本发明实施例提供一种众包项目中选取众包参与人的方法,包括:

[0009] 根据多个众包参与人分别提供的各自的行为记录和信用记录,获取所述多个众包参与人中每个众包参与人的信用信息和能力信息;

[0010] 从众包项目平台中确定需要分配参与人的众包项目,并从所述众包项目中获取到

项目信息,所述项目信息包括:所述众包项目的项目任务以及所述项目任务的内容信息;

[0011] 根据所述每个众包参与人的信用信息和能力信息为所述众包项目选取与所述项目信息匹配的众包参与人作为候选人集合,所述候选人集合包括:从所述多个众包参与人中筛选出的用于执行所述项目任务的众包参与人。

[0012] 第二方面,本发明实施例还提供一种众包项目中选取众包参与人的装置,包括:

[0013] 参与人信息处理模块,用于根据多个众包参与人分别提供的各自的行为记录和信用记录,获取所述多个众包参与人中每个众包参与人的信用信息和能力信息;

[0014] 众包项目信息处理模块,用于从众包项目平台中确定需要分配参与人的众包项目,并从所述众包项目中获取到项目信息,所述项目信息包括:所述众包项目的项目任务以及所述项目任务的内容信息;

[0015] 众包参与人选取模块,用于根据所述每个众包参与人的信用信息和能力信息为所述众包项目选取与所述项目信息匹配的众包参与人作为候选人集合,所述候选人集合包括:从所述多个众包参与人中筛选出的用于执行所述项目任务的众包参与人。

[0016] 从以上技术方案可以看出,本发明实施例具有以下优点:

[0017] 在本发明实施例中,根据多个众包参与人分别提供的各自的行为记录和信用记录,获取多个众包参与人中每个众包参与人的信用信息和能力信息,从众包项目平台中确定需要分配参与人的众包项目,并从众包项目中获取到项目信息,项目信息包括:众包项目的项目任务以及项目任务的内容信息,最后可以根据每个众包参与人的信用信息和能力信息为众包项目选取与项目信息匹配的众包参与人作为候选人集合,候选人集合包括:从多个众包参与人中筛选出的用于执行项目任务的众包参与人。本发明实施例中可以根据众包参与人的行为记录和信用记录确定出该参与人的信用信息和能力信息,从而保证对众包参与人的真实刻画,保证众包参与人存在欺诈的可能性足够小,并可以准确地描述出参与人的能力,对于需要分配参与人的众包项目,能够对该众包项目的项目任务做出解析,从而确定出项目任务的内容信息,该内容信息可以反映出该项目任务对参与人的要求,最后再通过前述步骤获取到的众包参与人的信用信息和能力信息,与项目任务进行匹配,筛选出的候选人集合中的众包参与人可以执行该项目任务,从而实现为众包项目有效选取参与人,保证筛选的可靠性,这与现有技术中依靠金钱担保选取参与人的方式是完全不同的,本发明实施例可以提升众包项目平台中参与人选取的效率。

附图说明

[0018] 为了更清楚地说明本发明实施例中的技术方案,下面将对实施例描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域的技术人员来讲,还可以根据这些附图获得其他的附图。

[0019] 图1为本发明实施例提供的一种众包项目中选取众包参与人的方法的流程方框示意图;

[0020] 图2为本发明实施例提供的众包项目中选取众包参与人的方法的系统架构示意图;

[0021] 图3为本发明实施例提供的众包项目中选取众包参与人的方法的一种应用场景下参与人选取的流程示意图;

[0022] 图4-a为本发明实施例提供的一种众包项目中选取众包参与人的装置的组成结构示意图；

[0023] 图4-b为本发明实施例提供的另一种众包项目中选取众包参与人的装置的组成结构示意图；

[0024] 图4-c为本发明实施例提供的一种参与人信息处理模块的组成结构示意图；

[0025] 图4-d为本发明实施例提供的一种众包项目信息处理模块的组成结构示意图；

[0026] 图4-e为本发明实施例提供的一种众包参与人选取模块的组成结构示意图；

[0027] 图5为本发明实施例提供的众包项目中选取众包参与人的方法应用于服务器的组成结构示意图。

具体实施方式

[0028] 本发明实施例提供了一种众包项目中选取众包参与人的方法和装置,用于为众包项目平台中的众包项目选取众包参与人,并能够保证筛选的可靠性。

[0029] 为使得本发明的发明目的、特征、优点能够更加的明显和易懂,下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,下面所描述的实施例仅仅是本发明一部分实施例,而非全部实施例。基于本发明中的实施例,本领域的技术人员所获得的所有其他实施例,都属于本发明保护的范围。

[0030] 本发明的说明书和权利要求书及上述附图中的术语“包括”和“具有”以及他们的任何变形,意图在于覆盖不排他的包含,以便包含一系列单元的过程、方法、系统、产品或设备不必限于那些单元,而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它单元。

[0031] 以下分别进行详细说明。

[0032] 本发明众包项目中选取众包参与人的方法的一个实施例,具体可以应用于为众包项目智能选取适合的参与人,请参阅图1所示,本发明一个实施例提供的众包项目中选取众包参与人的方法,可以包括如下步骤:

[0033] 101、根据多个众包参与人分别提供的各自的行为记录和信用记录,获取多个众包参与人中每个众包参与人的信用信息和能力信息。

[0034] 在本发明实施例中,众包也称为群众外包(英语:crowdsourcing)是互联网带来的新的生产组织形式。问题以公开招标的方式传播给未知的解决方案提供者群体。用户(这里指众包里的“众”)典型地组成在线社区并提交方案。群“众”还要审查方案,发现最好的。这些最好的方案最后由最先提出问题的一方所有,即归众包人(crowdsourcer)所有,并且群“众”中胜出的个人有时会被奖励。有时,这些工作会有不错的报酬,无论是金钱上的、奖励上的或者只是名声和知名度。另外一些情况,胜出者会有智力上的满足感。众包可以通过业余人士或志愿者利用他们的空余时间提供解决方案,或者让专家或小型企业从无人知晓到初具规模。常见的众包项目包括百科网站、问答类网站、开源项目、众包任务、众包标注、众包配送、论文评审等。

[0035] 在本发明实施例中的众包参与人指的是参与众包项目的参与人,在步骤101中多个众包参与人指的是准备用于筛选众包参与人的众包参与人集合,本发明实施例中对于众包参与人的筛选需要获取每个众包参与人的多方面信息,以期能够从多个角度尽可能的立

体地刻画出参与人的角色,具体的,对于多个众包参与人,需要先收集每个众包参与人的各自的行为记录和信用记录。其中,众包参与人的行为记录指的是该参与人已经产生的行为描述,例如该参与人已经执行过的众包项目以及完成情况,该参与人的社交记录、该参与人的学历情况、获奖情况等等,具体可以根据应用场景选取参与人的多种行为记录。众包参与人的信用记录是指该参与人在社会活动中的信用情况,这些信用记录可以来源于金融信用征集机构、司法机关、政府公共服务部门等。

[0036] 在本发明实施例中,通过收集众包参与人的行为记录和信用记录,可以尽量真实的客观的还原出众包参与人的实际情况,从而可以通过量化和标签的方式来描述每个众包参与人,例如可以获取到每个众包参与人的信用信息和能力信息,能力信息表示众包参与人的能力类型以及等级,信用信息可以表示众包参与人的信用水平等级。本发明实施例对于每个众包参与人,可以根据不同的参与人标识对每个众包参与人的能力信息和信用信息进行收集。

[0037] 举例说明,在对参与人信息处理中,需要对众包项目参与者的信用信息和能力信息进行收集、处理、存储,以供其他模块对于用户信用信息和能力的请求。这些信息包括用户提供的社交网络账号的相关行为,用户曾经参与的众包项目的完成情况和任务,用户社会信用信息,以及用户实名信息的收集,收集到的信息汇总存储于参与人信息存储数据库中。

[0038] 在本发明的一些实施例中,步骤101根据多个众包参与人分别提供的各自的行为记录和信用记录,获取多个众包参与人中每个众包参与人的信用信息和能力信息,包括:

[0039] A1、收集多个众包参与人的各自的用户信息,用户信息包括:历史众包行为记录、社交网络行为记录、社会信用记录和个人认证信息,其中,历史众包行为记录通过多个众包参与人在众包项目平台上的各自账户得到,社交网络行为记录通过多个众包参与人各自的社交网络账户得到,社会信用记录通过多个众包参与人各自的社会信用的授权查询得到,个人认证信息通过多个众包参与人各自提供的个人认证资料得到;

[0040] A2、根据多个众包参与人的各自的历史众包行为记录、社交网络行为记录和个人认证信息计算每个众包参与人的个人通用素质能力和专业知识能力;

[0041] A3、根据多个众包参与人的各自的社会信用记录和个人认证信息计算每个众包参与人的信用水平指标。

[0042] 具体的,在步骤A1中,众包参与人相关信息收集可以分为以下四个方面:众包平台信息收集、社交网络信息收集、社会信用查询授权和个人认证信息上传。例如对于众包平台信息收集,从各个可信任的众包平台对参与人的众包行为信息进行收集和整理,收集的过程需要用户主动提供其在各个众包平台上的相应账户信息。需要收集信息包括但不限于用户曾经参与的众包项目内容,参与方式和时长,参与角色,项目的进展和完成状况,做出的贡献大小,以及项目相关人员对于其工作的评价。对于社交网络信息收集,通过用户提供的社交网络账号和授权,利用网络爬虫从社交网络中获得其行为信息和好友关系,包括但不限于其发布的文字和图片状态,发布的地理位置信息,好友的评论和点赞信息,分享的内容和媒体,参加的社团活动信息。对于社会信用查询授权,通过用户对于其自身的社会信用提供查询的相关授权,信用授权获取模块将从获得用户授权的信用平台和管理机构自动获取其信用记录和信用评级,对于个人认证信息上传,要求用户上传其能够证明个人身份以及

其相关信息当中的社会身份的证明材料,可能需要上传的材料包括用户照片、合法身份证件信息、学历证明、职业证明、专业资格证明、相关项目和作品的创作者公证、社会关系证明、其他能够证明用户信用身份的证明材料。

[0043] 在步骤A2中,可以对收集到的用户信息进行解析,首先进行信息的解析与量化:由于原始信息的数据源众多,数据种类和形式复杂,因此需要将其转化成一致量化的、便于系统处理的数据格式。信息解析与量化加工的方式主要采用统计量化和文本化标签的方式进行。对于多媒体数据需要采用模式识别的方法从中提取关键信息。文本标签化的方法采用常见的自然语言处理方法进行,即通过利用预先训练好的分词器对这些文本信息进行分词处理,然后使用完整定义的自然语言处理流程进行分析,例如对于认证信息,主要需要解析其中能够判断或间接判断身份及能力的部分,例如身份证号、证书的类型、等级信息等,将其与一些可靠信息进行交叉(例如身份证号对应的出生省份,证书类型对应的学科和专业领域),可以进行标签化。

[0044] 在步骤A2和步骤A3中,从解析得到的数据向量中计算用户的各项能力值,以及众包信用值。对于用户的能力评价,主要集中于其个人通用素质能力(包括但不限于执行力、时间管理能力、沟通能力、创造力、组织能力)和专业水平(例如美术设计、文书策划、编程能力)的评估两方面,评估模型主要采用预先设置的能力模型和专家系统进行。专家系统可以看作是一类具有专门知识和经验的计算机智能程序系统,一般采用人工智能中的知识表示和知识推理技术来模拟通常由领域专家才能解决的复杂问题。对于用户信用水平,主要考虑其社会信用评级信息,以及认证信息与其他用户信息的一致性。

[0045] 在上述实现场景下,可以将众包参与人的能力信息划分为个人通用素质能力和专业知识能力,可以通过量化和分类的方式来描述参与人具有什么类型的能力以及具有什么等级的能力,对于众包参与人的信用信息可以通过信用水平指标来确定,信用水平指标描述了参与人的信用等级,该信用水平指标可以刻画出参与人的信用情况。

[0046] 进一步的,在本发明的另一些实施例中,步骤A3根据多个众包参与人的各自的社会信用记录和个人认证信息计算每个众包参与人的信用水平指标之后,本发明实施例提供的方法还包括如下步骤:

[0047] A4、计算多个众包参与人的各自的个人认证信息与收集到的用户信息之间的一致性指标,并保存得到的一致性计算结果。

[0048] 其中,众包参与人的各自的个人认证信息与用户信息是否一致,也能够从另一个侧面刻画出参与人的信用情况,认证信息与其他用户信息的一致性计算结果中,对于不一致的部分,留存在信用信息中,还可以用于反欺诈识别,详见后续实施例中的描述。

[0049] 102、从众包项目平台中确定需要分配参与人的众包项目,并从众包项目中获取到项目信息。

[0050] 其中,项目信息包括:众包项目的项目任务以及项目任务的内容信息。

[0051] 在本发明实施例中,众包项目平台中对于正在进行或者即将开展的众包项目,可以从该众包项目中获取到项目信息,众包项目由项目提供方提供,通过分析该众包项目的项目信息,可以客观描述出项目任务的内容信息。需要说明的是,对于不同类型的众包项目,需要收集的项目任务的内容信息可以是不同的,具体可以结合应用场景来确定。通常情况下,在众包项目信息处理中,可以收集众包项目的相关说明和内容,并从中分析得到项目

的任务信息、范围、时间、费用、质量、资源、风险等核心信息,以用于选取众包参与人使用。生成的信息存储于众包项目信息存储数据库中。其中,范围指的是任务的知识领域涉及的范围,资源是指任务发起人能够提供的除金钱以外的支持,风险是指任务失败或者不能按时完成可能造成的损失和惩罚。

[0052] 在本发明实施例中,对于项目任务的内容解析可以通过自然语言处理和模板匹配进行提取,即通过自然语言处理进行模糊匹配,查找相关关键词和相近词,然后将相关信息提取量化。其中,自然语言生成系统把计算机数据转化为自然语言,自然语言理解系统把自然语言转化为计算机程序更易于处理的形式。主要包括自动分词、词性标注、句法分析、文本分类、信息抽取、语义理解等多个模块构成。主题模型分析是指主题模型(Topic Model)在机器学习和自然语言处理等领域是用来在一系列文档中发现抽象主题的一种统计模型。主题模型自动分析每个项目任务,统计项目任务中的词语,根据统计的信息来断定当前项目任务含有哪些主题,以及每个主题所占的比例各为多少。

[0053] 在本发明的一些实施例中,步骤102从众包项目中获取到项目信息,包括:

[0054] B1、确定众包项目的项目任务,获取项目任务的内容信息,内容信息包括:众包项目的名称、类型、知识领域范围、时间、费用、质量、资源、风险;

[0055] B2、根据项目任务的内容信息估计项目任务所需的成本和项目任务的完成要求,其中,项目任务所需的成本包括:人力成本、资金成本和时间成本,人力成本包括:项目任务所需的专业技能类型、人员数量和专业技能水平,资金成本包括:项目任务的项目报酬、项目沟通渠道所需要的资金、项目资源提供成本、项目开销成本,时间成本包括:项目沟通成本和项目工时成本,完成要求包括:完成可靠度、完成质量、未完成和延期的损失。

[0056] 具体的,在步骤B1中,众包项目信息获取是指众包平台上正在进行或即将开展的众包项目进行相关描述信息和任务内容信息的收集。主要包括众包项目的名称、类型,众包项目的描述信息,以及完成项目的期限和报酬。接下来进行步骤B2,对项目信息进行分析,将收集到的众包项目信息进行解析,从中给出项目资源和进度管理所需要的信息,主要包括项目所需的人力和资金成本估计,时间成本估计,以及完成要求说明。其中人力成本估计应当以所需的专业技能的人力单位计算,对于资金成本,应当计算其中显性和隐性的各项现金与物资开销。对于时间成本,主要在于沟通成本和工时成本。完成要求是指该众包项目对于最终结果完成情况的要求,具体包括完成可靠度,完成质量,未完成和延期的损失等。对于以上信息的提取,采用与前文所述用户信息的解析与量化过程类似的方法,采用统计量化和文本标签化的方法进行,并结合模式识别和专家系统对数据中的深层信息进行分析和提取。在项目进行过程中,通过对于项目的进度反馈和评估还将追加更多的信息作为分析依据,不断迭代调整这些数据。

[0057] 需要说明的是,在前述实施例中,步骤101和步骤102之间没有先后逻辑顺序,可以先执行步骤101再执行步骤102,也可以先执行步骤102再执行步骤101,还可以同时执行步骤101和步骤102,其中,图1中以先执行步骤101再执行步骤102为例进行说明,本发明实施例也可以采用其它顺序步骤,此处不做限定。

[0058] 103、根据每个众包参与人的信用信息和能力信息为众包项目选取与项目信息匹配的众包参与人作为候选人集合。

[0059] 其中,候选人集合包括:从多个众包参与人中筛选出的用于执行项目任务的众包

参与人。

[0060] 在本发明实施例中,通过前述步骤101和步骤102获取到众包参与人的信用信息和能力信息、众包项目的项目信息之后,再通过这两者的匹配,就可以从待选的多个众包参与人中筛选出候选人集合,其中,众包参与人的信用信息和能力信息刻画出的众包参与人的个人情况,众包项目的项目信息刻画出的项目任务的任务情况,则众包参与人的个人情况能够适应于项目任务的任务情况需求就可以确定两者匹配,具体的匹配方式可以有多种,例如,对于众包参与人选取,利用参与人信息存储数据库和众包项目信息存储数据库中的相关信息,为众包项目匹配合适的参与人,其中包含了对于项目和参与人双方的特性分析,还可以包括反欺诈校验的判断,分别由参与人匹配模块和反欺诈校验模块完成,参与人的选取过程可以由相关控制策略控制完成,详见后续实施例中的举例说明。

[0061] 在本发明的一些实施例中,在前述执行步骤B1至步骤B2的实现场景下,步骤103根据每个众包参与人的信用信息和能力信息为众包项目选取与项目信息匹配的众包参与人作为候选人集合,包括:

[0062] C1、根据每个众包参与人的信用信息和能力信息、项目任务的项目信息,预测每个众包参与人的项目接受可能性指标、项目中途放弃可能性指标;

[0063] C2、根据每个众包参与人的能力信息和每个众包参与人的项目接受可能性指标,从多个众包参与人中查找到符合项目任务所需的人力成本、且符合项目任务的完成质量的第一众包参与人子集;

[0064] C3、根据每个众包参与人的信用信息从多个众包参与人中查找到符合项目任务的完成可靠度的第二众包参与人子集;

[0065] C4、根据每个众包参与人的能力信息和每个众包参与人的项目中途放弃可能性指标,从多个众包参与人中查找到符合项目任务所需的时间成本、资金成本、且符合项目任务的未完成和延期的损失的第三众包参与人子集;

[0066] C5、根据第一众包参与人子集、第二众包参与人子集、第三众包参与人子集确定出与项目信息匹配的众包参与人作为候选人集合。

[0067] 具体的,众包项目参与人匹配中,对于众包项目的参与人选取和匹配,可以通过步骤C1至步骤C4的多个方面来确定,接下来进行举例说明,在步骤C1中,首先进行用户意愿分析,需要分析各个用户对于项目的意愿水平,即用户接受该任务的可能性,以及用户中途放弃该任务的可能性。例如对于用户意愿的分析主要考虑项目的最终报酬,以及用户曾经的众包行为记录,可以通过简单的概率预测模型得到。在步骤C2中进行人力需求与参与人能力匹配,主要通过参与人的能力与项目人力需求之间的人力资源匹配,保证找到的众包项目参与人拥有能够完成该项目的的能力,并保证这一项目能够达到预期的质量。在步骤C3中进行完成可靠度与参与人信用匹配,通过找到足够信用水平的参与人来保证项目的完成可靠度,在步骤C4中进行项目进度保证,除了步骤C2和步骤C3之外,还需要考虑项目的整体进度与规划情况,主要涉及时间成本和资金成本,以及可能的风险。在满足以上条件之后,可确定选择的候选人集合。例如,通过前述步骤C2至步骤C4分别确定出第一众包参与人子集、第二众包参与人子集、第三众包参与人子集,则这三个众包参与人子集取交集,就可以确定出最终的候选人集合,即要求候选人集合中筛选出的众包参与人能够同时满足步骤C2至步骤C4中的多个方面要求,从而保证众包任务的顺利进行。

[0068] 进一步的,在本发明的一些实施例中,前述的步骤C3根据每个众包参与人的信用信息从多个众包参与人中查找到符合项目任务的完成可靠度的第二众包参与人子集,包括:

[0069] C31、若一致性计算结果表示众包参与人的个人认证信息与收集到的用户信息存在不一致时,将一致性计算结果存储为反欺诈识别信息;

[0070] C32、根据每个众包参与人的信用信息和反欺诈识别信息从多个众包参与人中查找到符合项目任务的完成可靠度的第二众包参与人子集。

[0071] 其中,反欺诈识别可以通过反欺诈策略给出评分,例如对于用户的上报年龄和证件上登记的年龄不一致时,将对用户的信用分数扣除预设的值,以显示其中的风险。反欺诈策略使用预先配置的方式确定。

[0072] 在本发明的一些实施例中,步骤103根据每个众包参与人的信用信息和能力信息为众包项目选取与项目信息匹配的众包参与人作为候选人集合之后,本发明实施例中提供的方法还包括:

[0073] D1、将选取的候选人集合通知给众包项目平台,由众包项目平台根据候选人集合对项目任务进行分发;

[0074] D2、根据候选人集合中的各个众包参与人的工作进度对项目任务的项目进度进行评估,根据评估结果预测项目任务的项目进度是否符合预期,再根据预测结果调整候选人集合。

[0075] 其中,众包项目跟踪过程中,需要跟踪和评估众包项目的进展和完成效果,并反馈给众包参与人选取子系统作为调整匹配参与人的依据,同时也会同步到众包项目信息存储数据库中进行留存。对于众包项目进度预估,在完成候选人集合的选择和任务的分发之后,众包项目的进度将随着众包候选人的工作推进,在此之后,需要系统不断监控各个用户的工作进度,并对其中已完成的工作量进行评估,并根据项目已经进行的时间、进度和原先预估的时间成本预估当前的项目的进度是否符合预期,并反馈给项目信息分析模块进行信息的调整。调整项目信息分析得到的结果,即对于人力单位、能力要求等进行调整,接着再根据这些信息对候选人进行调整。

[0076] 通过以上实施例对本发明实施例的描述可知,根据多个众包参与人分别提供的各自的行为记录和信用记录,获取多个众包参与人中每个众包参与人的信用信息和能力信息,从众包项目平台中确定需要分配参与人的众包项目,并从众包项目中获取到项目信息,项目信息包括:众包项目的项目任务以及项目任务的内容信息,最后可以根据每个众包参与人的信用信息和能力信息为众包项目选取与项目信息匹配的众包参与人作为候选人集合,候选人集合包括:从多个众包参与人中筛选出的用于执行项目任务的众包参与人。本发明实施例中可以根据众包参与人的行为记录和信用记录确定出该参与人的信用信息和能力信息,从而保证对众包参与人的真实刻画,保证众包参与人存在欺诈的可能性足够小,并可以准确地描述出参与人的能力,对于需要分配参与人的众包项目,能够对该众包项目的项目任务做出解析,从而确定出项目任务的内容信息,该内容信息可以反映出该项目任务对参与人的要求,最后再通过前述步骤获取到的众包参与人的信用信息和能力信息,与项目任务进行匹配,筛选出的候选人集合中的众包参与人可以执行该项目任务,从而实现为众包项目有效选取参与人,保证筛选的可靠性,这与现有技术中依靠金钱担保选取参与人

的方式是完全不同的,本发明实施例可以提升众包项目平台中参与人选取的效率。

[0077] 为便于更好的理解和实施本发明实施例的上述方案,下面举例相应的应用场景来进行具体说明。

[0078] 本发明实施例利用众包参与人的多方面信息,包括其过往的众包行为记录、其社交网站的好友记录、以及其他社会行为的信用记录等,从候选的众包对象中选择恰当数量的可靠的特定参与人,并且能够保证在这些参与人的参与下,该众包结果能够在要求的水平上可信赖,从而保证众包行为的质量达到要求,提升众包行为的效率。本发明实施例提供的方法的系统结构图如图2所示,可以由四个部分组成,分别是参与人信息处理、众包项目信息处理、众包参与人选取、众包项目跟踪。本发明能够有效收集和处理众包参与人的相关信息,动态生成参与人相关能力和信用的评估结果,并结合众包项目的需求,对参与人和项目进行匹配筛选,为众包项目找到合适的参与对象,并且全程跟踪这些参与对象和项目之间的进度,保证众包项目能够有效顺利地推进。下面将详细介绍该发明技术方案的实现过程。

[0079] 接下来对于图2所示的对于各个部分的构成进行如下举例说明。

[0080] 首先,对于参与人信息处理部分说明如下,首先进行众包项目参与者的信用信息收集、处理、存储,以及处理其他模块对于用户信用信息的请求。这些信息包括用户提供的社交网络账号的相关行为,用户曾经参与的众包项目的完成情况和任务,用户信用信息,以及用户实名信息的收集,分别由图中的社交网络爬虫、众包行为信息模块、信用授权获取模块、实名信息上传收集模块控制,收集到的信息汇总存储于参与人信息存储数据库中。

[0081] 对于众包项目信息处理部分说明如下,收集众包项目的相关说明和内容,并从中分析得到项目的任务信息、范围、时间、费用、质量、资源、风险等核心信息,以用于选取众包参与人使用。生成的信息存储于众包项目信息存储数据库中。

[0082] 对于众包参与人选取部分说明如下,利用参与人信息存储数据库和众包项目信息存储数据库中的相关信息,为众包项目匹配合适的参与人。其中包含了对于项目和参与人双方的特性分析和反欺诈校验,分别由参与人匹配模块和反欺诈校验模块完成。参与人的选取过程可以由相关控制策略模块控制。

[0083] 最后对于众包项目跟踪说明如下,用于跟踪和评估众包项目的进展和完成效果,并反馈给众包参与人选取子系统作为调整匹配参与人的依据,同时也会同步到众包项目信息存储数据库中进行留存。

[0084] 其中,各个子系统和子模块的构成不局限于上文所描述的结构,还包含省略任意模块或子模块、重新拆解组合各个子模块构成新模块的系统结构。

[0085] 接下来请参阅图3所示,本发明对于众包项目和参与的信息收集、处理、匹配、追踪的相关流程图2众包参与人选取流程,对于流程的详细说明如下:

[0086] 11、众包参与人相关信息收集:参与人信息收集主要分为以下四个方面:

[0087] 1)、众包平台信息收集:从各个可信任的众包平台对参与人的众包行为信息进行收集和整理,收集的过程需要用户主动提供其在各个众包平台上的相应账户信息。需要收集信息包括但不限于用户曾经参与的众包项目内容,参与方式和时长,参与角色,项目的进展和完成状况,做出的贡献大小,以及项目相关人员对于其工作的评价。该步骤由参与人信

息处理子系统众包行为信息模块完成,记作 $\{U_p^{(id)}\}$,其中上标id表示用户编号。该部分信息需要直接存储到参与人信息存储数据库中。

[0088] 2)、社交网络信息收集:通过用户提供的社交网络账号和授权,利用网络爬虫从社交网络中获得其行为信息和好友关系,包括但不限于其发布的文字和图片状态,发布的地理位置信息,好友的评论和点赞信息,分享的内容和媒体,参加的社团活动信息。该步骤由参与人信息处理子系统中的社交网络爬虫完成,获得的信息记作 $\{U_s^{(id)}\} = \{U_{ss}^{(id)}, U_{sf}^{(id,1)}, \dots, U_{sf}^{(id,m)}\}$,其中 $U_{ss}^{(id)}$ 表示该用户的行为, $U_{sf}^{(id,j)}$ 表示其第j个好友与其交互的行为。

[0089] 其中,网络爬虫是一种“自动化浏览网络”的程序,或者说是一种网络机器人。它们被广泛用于互联网搜索引擎或其他类似网站,以获取或更新这些网站的内容和检索方式。它们可以自动采集所有其能够访问到的页面内容,以供搜索引擎做进一步处理(分检整理下载的页面),而使得用户能更快的检索到他们需要的信息。

[0090] 3)、社会信用查询授权:通过用户对于其自身的社会信用提供查询的相关授权,信用授权获取模块将从获得用户授权的信用平台和管理机构自动获取其信用记录和信用评级,记作 $\{U_c^{(id)}\}$ 。其中 $U_c^{(id)}$ 包含了查询的平台、用户身份、信用记录和信用评级、查询时间、有效时间等信息。常见的社会信用机构包括金融信用征集机构、司法机关、政府公共服务部门等。

[0091] 4)、个人认证信息上传:要求用户上传其能够证明个人身份及其相关信息当中的社会身份的证明材料,可能需要上传的材料包括用户照片、合法身份证件信息、学历证明、职业证明、专业资格证明、相关项目和作品的创作者公证、社会关系证明、其他能够证明用户信用身份的证明材料。要求的内容包括其中的一项或多项,具体情况根据用户提供的其他信息来决定。收集到的认证信息记作 $\{U_a^{(id)}\}$ 。

[0092] 从这些数据收集流程中得到的数据均以用户的唯一标识编号id作为索引,以方便后续的流程查询和使用,记作 $\{U^{(id)}\} = \{U_p^{(id)}, U_s^{(id)}, U_c^{(id)}, U_a^{(id)}\}$ 。

[0093] 12、用户信息分析:用户信息分析可以细分为3个小步骤,分别是信息的解析与量化,分析用户信用与能力,结果的结构化存储。下面进行详细的介绍:

[0094] 1)、信息的解析与量化:由于原始信息的数据源众多,数据种类和形式复杂,因此需要将其转化成一致量化的、便于系统处理的数据格式。信息解析与量化加工的方式主要采用统计量化和文本化标签的方式进行,例如对于用户发布状态的次数、天数、点赞数量、评论数量和时间点信息等进行统计,或者对用户发送的内容、评论内容、地理位置标记等进行文本标签化。对于多媒体数据(主要是用户的头像、照片、证件照等,或者拍摄的用于证明其身份的视频、语音等)需要采用模式识别的方法从中提取关键信息,例如使用人脸识别技术对证件和账户头像信息进行识别和比较,使用OCR方法对图像和视频中的证件信息进行识别等。文本标签化的方法采用常见的自然语言处理方法进行,即通过利用预先训练好的分词器对这些文本信息进行分词处理,然后使用完整定义的自然语言处理流程进行分析,例如先分析其中的词义并生成0-1向量表示,然后使用主题模型分析其中的主题,并使用人

工标注的方法对提取的主题进行标签化。这里可以使用例如词向量、文本向量嵌入的方法，叠加聚类的方法，或者LDA (Latent Dirichlet Allocation) 概率主题方法，以及深度神经网络等主流自然语言处理方法以及他们的组合来完成，也可以借助第三方计算平台提供的自然语言处理云服务完成 (例如自然语言处理服务)。对于认证信息 $\{U_a^{(i)}\}$ ，主要需要解析

其中能够判断或间接判断身份及能力的部分，例如身份证号、证书的类型、等级信息等，将其与一些可靠信息进行交叉 (例如身份证号对应的出生省份，证书类型对应的学科和专业领域)，可以进行标签化。解析得到的数据可以表示为一个包含多个维度的向量，向量的每一维度存储着某一方面的数据值，标签化的列则表示用户拥有这个标签的可能性。

[0095] 其中，在主题模型分析中，主题模型 (Topic Model) 在机器学习和自然语言处理等领域是用来在一系列文档中发现抽象主题的一种统计模型。直观来讲，如果一篇文章有一个中心思想，那么一些特定词语会更频繁的出现。比方说，如果一篇文章是在讲狗的，那“狗”和“骨头”等词出现的频率会高些。如果一篇文章是在讲猫的，那“猫”和“鱼”等词出现的频率会高些。而有些词例如“这个”、“和”大概在两篇文章中出现的频率会大致相等。但真实的情况是，一篇文章通常包含多种主题，而且每个主题所占比例各不相同。因此，如果一篇文章10%和猫有关，90%和狗有关，那么和狗相关的关键字出现的次数大概是和猫相关的关键字出现次数的9倍。一个主题模型试图用数学框架来体现文档的这种特点。主题模型自动分析每个文档，统计文档内的词语，根据统计的信息来断定当前文档含有哪些主题，以及每个主题所占的比例各为多少。在自然语言处理中，自然语言生成系统把计算机数据转化为自然语言。自然语言理解系统把自然语言转化为计算机程序更易于处理的形式。主要包括自动分词、词性标注、句法分析、文本分类、信息抽取、语义理解等多个模块构成。在模式识别中，通过计算机用数学技术方法来研究模式的自动处理和判读。把环境与客体统称为“模式”。

[0096] 2)、分析用户信用和能力：从解析得到的数据向量中计算用户的各项能力值，以及众包信用值。对于用户的能力评价，主要集中于其个人通用素质能力 (包括但不限于执行力、时间管理能力、沟通能力、创造力、组织能力) 和专业知识水平 (例如美术设计、文书策划、编程能力) 的评估两方面，评估模型主要采用预先设置的能力模型和专家系统进行，例如对于沟通能力，专家系统会从其社交网络行为 $\{U_s^{(i)}\}$ 、历史众包项目信息 $\{U_p^{(i)}\}$ 判断其能力的级别。得到的能力指标可表示为：

$$[0097] \quad \{ability^{(i)}\} = \left\{ \left\langle A_{person}^{(i)}, A_{profession}^{(i)} \right\rangle \right\} = \left\{ \left\{ \left\langle a_p^{(i)} \right\rangle, \left\langle a_{pro}^{(i)} \right\rangle \right\} \right\}。$$

[0098] 对于用户信用水平，主要考虑其社会信用评级信息 $\{U_c^{(i)}\}$ ，以及认证信息 $\{U_a^{(i)}\}$ 与其他信息的一致性，对于不一致的部分，留存在信用信息中，用于反欺诈识别。具体的信用水平可以通过以下公式进行计算：

$$[0099] \quad credit^{(i)} = \alpha \sum_{u_c^{(i)} \in U_c^{(i)}} nom(u_c^{(i)}) + \beta \sum_{\substack{u_a^{(i)} \in U_a^{(i)} \\ u^{(i)} \notin U_a^{(i)}}} consist(u_a^{(i)}, u^{(i)}) + \gamma \|U^{(i)}\|；$$

[0100] 其中,公式中 α 、 β 、 γ 代表三个权值,表示三部分的重要程度,通过人工调整确定,所有用户采用统一的权值设定, $u_c^{(id)} \in U_c^{(id)}$ 、 $\text{norm}(U_c^{(id)})$,代表前文提到的单个用户的社会信用查询结果的每一项记录进行标准化的值,

$u_a^{(id)} \in U_a^{(id)}$ 、 $u_c^{(id)} \notin U_a^{(id)}$ 、 $\text{consist}(U_a^{(id)}, U_c^{(id)})$ 、代表指对用户的各项认证信息中的每一项和其他信息中的每一项进行一致性校验,通过一致性校验函数找出其中能够互相印证的信息(例如姓名、工作等),并作为信用衡量的因素。 $\|U^{(id)}\|$ 表示该用户所有信息条目的数量,数量越大认为越可靠,即用户的信用通过标准化的用户信用分数、用户认证信息与其他信息的一致性、用户信息的完整度三个部分加权得到。

[0101] 3)、结构化存储:对于能力指标和信用信息包含了各个方面的能力和信用信息,以及不一致的留存信息,需要通过结构化的存储才能存储于数据库中以便于检索。为此,可以定义一套能力指标的分类索引和信用信息的列表索引,对以上结果进行存储。例如对于专业能力指标 $\{a_{pro}^{(id)}\}$,可以按照一级行业,二级行业进行类目划分,然后把每个能力放置于相应的类目之下。对于信用信息,逐一列出信用指标 $\text{credit}^{(id)}$,以及影响信用指标的 $\text{consist}(U_a^{(id)}, U_c^{(id)})$ 项目。

[0102] 13、众包项目信息获取:对众包平台上正在进行或即将开展的众包项目进行相关描述信息和任务内容信息的收集。主要包括众包项目的名称、类型,众包项目的描述信息,以及完成项目的期限和报酬。这些信息可以是图片、文本、视频、音频等各种媒体,由众包平台提供。

[0103] 14、项目信息分析:将收集到的众包项目信息进行解析,从中给出项目资源和进度管理所需要的信息,主要包括项目所需的人力和资金成本估计,时间成本估计,以及完成要求说明。其中人力成本估计应当以所需的专业技能的人力单位计算,即对于所需的每项专业技能都应当计算得到其所需要的不同能力和水平的人力单位,记作 $\{H^{(proj)}\} = \left\{ \left\langle H_{type}^{(proj)}, H_{num}^{(proj)}, H_{level}^{(proj)} \right\rangle \right\}$,其中 $proj$ 表示项目的索引,由 $H_{type}^{(proj)}$ 类型、 $H_{num}^{(proj)}$ 数量和 $H_{level}^{(proj)}$ 专业水平构成。对于资金成本,应当计算其中显性和隐性的各项现金与物资开销,即包括项目的最终报酬 $C_{reward}^{(proj)}$,项目沟通渠道所需要的资金 $C_{comm}^{(proj)}$,项目物料成本和资源提供成本 $C_{supp}^{(proj)}$,还有其他开销 $C_{other}^{(proj)}$,记作

$\{C^{(proj)}\} = \left\{ \left\langle C_{reward}^{(proj)}, C_{comm}^{(proj)}, C_{supp}^{(proj)}, C_{other}^{(proj)} \right\rangle \right\}$ 。对于时间成本,主要在于沟通成本 $T_{comm}^{(proj)}$ 和工时成本 $T_{work}^{(proj)}$,记作 $\{T^{(proj)}\} = \left\{ \left\langle T_{comm}^{(proj)}, T_{work}^{(proj)} \right\rangle \right\}$ 。完成要求是指该众包项目对于最终结果完成情况的要求,具体包括完成可靠度 $F_{reli}^{(proj)}$,完成质量 $F_{quali}^{(proj)}$,未完成和延期损失 $F_{risk}^{(proj)}$ 等,记作 $\{F^{(proj)}\} = \left\{ \left\langle F_{reli}^{(proj)}, F_{quali}^{(proj)}, F_{risk}^{(proj)} \right\rangle \right\}$ 。对于以上信息的提取,采用与前文所述用户

信息的解析与量化过程类似的方法,采用统计量化和文本标签化的方法进行,并结合模式识别和专家系统对数据中的深层信息进行分析和提取。由于在项目尚未开始时,信息来源仅为项目发起者提供的基本描述信息,因此这里的分析得到的数据并不一定能够反映该众包项目真实的情况。在项目进行过程中,通过对于项目的进度反馈和评估还将追加更多的信息作为分析依据,不断迭代调整这些数据。

[0104] 15、信息存储:把用户信息分析和项目信息分析得到的结果分布存储于参与人信息存储模块和项目信息存储模块中,并将不同参与人的信息使用用户id作为主键进行存储,将不同项目的信息使用项目编号proj作为主键存储,方便后续进行匹配查询使用。

[0105] 16、众包项目参与人匹配:对于众包项目的参与人选取和匹配,通过众包参与人选取子系统完成,主要包括了以下几个方面:

[0106] 1)、用户意愿分析:首先需要分析各个用户对于项目的意愿水平,即用户接受该任务的可能性 $\{P_{ac}^{(i),proj}\}$, 以及用户中途放弃该任务的可能性 $\{P_{ab}^{(i),proj}\}$ 。对于用户意愿的分析主要考虑项目的最终报酬 $C_{reward}^{(proj)}$, 以及用户曾

[0107] 经的众包行为记录 $\{U_p^{(i)}\}$, 可以通过简单的概率预测模型得到:

$$P_{ac}^{(i),proj} = P\left(\text{accept} \mid U_p^{(i)}, C_{reward}^{(proj)}\right), P_{ab}^{(i),proj} = P\left(\text{abandon} \mid U_p^{(i)}, C_{reward}^{(proj)}\right),$$
 其中

[0108] 的概率分布通过朴素贝叶斯方法学习得到。在机器学习中,朴素贝叶斯分类器是一系列以假设特征之间强(朴素)独立下运用贝叶斯定理为基础的简单概率分类器。

[0109] 2)、人力需求与参与人能力匹配:主要通过参与人的能力 $\{\text{ability}^{(i)}\}$ 与项目人力需求 $\{H^{(proj)}\}$ 之间的函数关系(即下文所定义的人力资源函数)保证找到的众包项目参与人拥有能够完成该项目的的能力,并保证这一项目能够达到预期的质量。这一过程可以表示为:

$$[0110] \quad H_{num}^{(proj)} \leq \lambda_1 \sum_{i \in S} h\left(\text{ability}^{(i)}, H_{type}^{(proj)}, H_{level}^{(proj)}, P_{ac}^{(i),proj}\right);$$

$$[0111] \quad F_{quali}^{(proj)} > \lambda_2 \sum_{i \in S} q\left(\text{ability}^{(i)}, H_{type}^{(proj)}, H_{level}^{(proj)}, P_{ac}^{(i),proj}\right);$$

[0112] 其中 λ_1 和 λ_2 为放大因子, S 为选择的索引集合, h 表示人力资源函数, 定义为:

$$[0113] \quad h\left(\text{ability}^{(i)}, H_{type}^{(proj)}, H_{level}^{(proj)}, P_{ac}^{(i),proj}\right) = P_{ac}^{(i),proj} \sum_{type} I\left(A_{profession}^{(i)} \cap H_{type}^{(proj)} \neq \emptyset\right) \left\| A_{profession}^{(i)} - H_{level}^{(proj)} \right\|;$$

[0114] 其中, I() 为指示函数, 即参与人能力专业能力匹配、专业水平满足要求, 再加上其个人能力乘以其接受意愿。

[0115] q 表示质量效果函数, 定义为:

$$[0116] \quad q\left(\text{ability}^{(i)}, H_{type}^{(proj)}, H_{level}^{(proj)}, P_{ac}^{(i),proj}\right) = P_{ac}^{(i),proj} \sum_{type} I\left(A_{profession}^{(i)} \cap H_{type}^{(proj)} \neq \emptyset\right) \left\| A_{person}^{(i)} \right\| \text{credit}^{(i)};$$

[0117] 其中, I() 为指示函数, 即个人基本能力和信用保证专业能力能够发挥的水平。

[0118] 3)、完成可靠度与参与人信用匹配:通过找到足够信用水平的参与人来保证项目的完成可靠度,可以表示为:

$$[0119] \quad F_{reli}^{(proj)} \leq \lambda_3 \sum_{i \in S} credit^{(i)} + antifraud(U_a^{(i)});$$

[0120] 其中, $antifraud(U_a^{(i)})$ 表示反欺诈模块对于用户个人认证信息的评价。反欺诈模块通过反欺诈策略给出这一评分, 例如对于用户的上报年龄和证件上登记的年龄不一致时, 将对用户的信用分数扣除预设的值, 以显示其中的风险。反欺诈策略使用人工设定的方法给出。

[0121] 4)、项目进度保证: 除了以上几点, 还需要考虑项目的整体进度与规划情况, 主要涉及时间成本 $\{T^{(proj)}\}$ 和资金成本 $\{C^{(proj)}\}$, 以及可能的风险 $F_{risk}^{(proj)}$ 。这些关系可以表示为:

$$[0122] \quad T_{comm}^{(proj)} + T_{work}^{(proj)} > \frac{\lambda_4 H_{num}^{(proj)}}{\sum_{i \in S} h(\text{ability}^{(i)}, H_{type}^{(proj)}, H_{level}^{(proj)}, P_{ac}^{(i, proj)})} + \|S\| \sum_{i \in S} P_{ab}^{(i, proj)};$$

$$[0123] \quad F_{risk}^{(proj)} > C_{reward}^{(proj)}.$$

[0124] 在满足以上条件之后, 可确定选择的候选人集合 S , 并通知众包平台进行任务分发, 将任务分发给选中的候选人。

[0125] 17、众包项目进度预估: 在完成候选人集合的选择和任务的分发之后, 众包项目的进度将随着众包候选人的工作推进, 在此之后, 需要系统不断监控各个用户的工作进度, 并对其中已完成的工作量 $H_{fin}^{(proj)}$ 进行评估, 并根据项目已经进行的时间、进度和原先预估的时间成本 $T^{(proj)}$ 预估当前的项目的进度是否符合预期, 并反馈给项目信息分析模块进行信息的调整。调整项目信息分析得到的结果, 即对于人力单位、能力要求等进行调整, 接着再根据这些信息对候选人进行调整。

[0126] 通过前述对本发明的举例说明可知, 本发明具有可信赖性, 本发明从多个角度收集众包参与人的信息, 尽可能立体地刻画出参与人的角色, 保证参与众包项目的人存在欺诈的可能性足够小, 并且结合项目相关的指标, 如项目完成可靠度对其进行量化, 能够实时发现和排除不可信任的参与人对于众包项目的潜在风险。本发明实施例可以保证众包项目质量, 本发明通过对项目构成和资源需求进行量化, 结合对于参与人用户的能力模型, 以及其历史的相关行为, 对用户的意愿和信用进行了评价, 在多个方面保证了项目的质量。这些提升主要体现在以下方面:

[0127] 1、从用户的个人能力和专业能力两个方面刻画用户的水平, 不但考虑完成某项任务的专业技能, 同时考虑在项目进行中的执行力、沟通能力等, 并与项目的相关成本和要求挂钩, 使得项目的进展能够顺利进行。

[0128] 2、对用户的意愿进行了评价, 使得在评估参与人的能力的同时还考虑了参与人数量可能不足的问题, 对于人力成本的控制有更加细致的手段。

[0129] 3、在分配完成众包项目的候选参与人集合后, 仍然不断跟踪项目的相关进度, 并及时根据评估的结果调整相关数值, 根据需要调整候选人集合的大小, 保证了项目的质量和进度。

[0130] 另外, 本发明实施例还可以提升众包平台的效率, 本发明通过合理的人员安排, 从

众包平台中选择最为合适的候选人群体,在保证项目的风险足够低的同时消耗了尽可能少的人力,既降低了项目的人力成本,同时也使得更多的候选人可以参与到更多的项目当中,从而有效提升了众包平台的效率。

[0131] 需要说明的是,对于前述的各方法实施例,为了简单描述,故将其都表述为一系列的动作组合,但是本领域技术人员应该知悉,本发明并不受所描述的动作顺序的限制,因为依据本发明,某些步骤可以采用其他顺序或者同时进行。其次,本领域技术人员也应该知悉,说明书中所描述的实施例均属于优选实施例,所涉及的动作和模块并不一定是本发明所必须的。

[0132] 为便于更好的实施本发明实施例的上述方案,下面还提供用于实施上述方案的相关装置。

[0133] 请参阅图4-a所示,本发明实施例提供的一种众包项目中选取众包参与人的装置400,可以包括:参与人信息处理模块401、众包项目信息处理模块402和众包参与人选取模块403,其中,

[0134] 参与人信息处理模块401,用于根据多个众包参与人分别提供的各自的行为记录和信用记录,获取所述多个众包参与人中每个众包参与人的信用信息和能力信息;

[0135] 众包项目信息处理模块402,用于从众包项目平台中确定需要分配参与人的众包项目,并从所述众包项目中获取到项目信息,所述项目信息包括:所述众包项目的项目任务以及所述项目任务的内容信息;

[0136] 众包参与人选取模块403,用于根据所述每个众包参与人的信用信息和能力信息为所述众包项目选取与所述项目信息匹配的众包参与人作为候选人集合,所述候选人集合包括:从所述多个众包参与人中筛选出的用于执行所述项目任务的众包参与人。

[0137] 在本发明的一些实施例中,请参阅图4-b所示,所述众包项目中选取众包参与人的装置400还包括:

[0138] 众包项目跟踪模块404,用于所述众包参与人选取模块403根据所述每个众包参与人的信用信息和能力信息为所述众包项目选取与所述项目信息匹配的众包参与人作为候选人集合之后,将选取的所述候选人集合通知给所述众包项目平台,由所述众包项目平台根据所述候选人集合对所述项目任务进行分发;根据所述候选人集合中的各个众包参与人的工作进度对所述项目任务的项目进度进行评估,根据评估结果预测所述项目任务的项目进度是否符合预期,再根据预测结果调整所述候选人集合。

[0139] 在本发明的一些实施例中,请参阅图4-c所示,所述参与人信息处理模块401,包括:

[0140] 收集模块4011,用于收集多个众包参与人的各自的用户信息,所述用户信息包括:历史众包行为记录、社交网络行为记录、社会信用记录和个人认证信息,其中,所述历史众包行为记录通过多个众包参与人在所述众包项目平台上的各自账户得到,所述社交网络行为记录通过多个众包参与人各自的社交网络账户得到,所述社会信用记录通过多个众包参与人各自的社会信用的授权查询得到,所述个人认证信息通过多个众包参与人各自提供的个人认证资料得到;

[0141] 能力计算模块4012,用于根据所述多个众包参与人的各自的历史众包行为记录、社交网络行为记录和个人认证信息计算每个众包参与人的个人通用素质能力和专业知识

能力；

[0142] 信用计算模块4013,用于根据所述多个众包参与人的各自的社会信用记录和个人认证信息计算每个众包参与人的信用水平指标。

[0143] 在本发明的一些实施例中,所述信用计算模块4013,还用于根据所述多个众包参与人的各自的社会信用记录和个人认证信息计算每个众包参与人的信用水平指标之后,计算所述多个众包参与人的各自的个人认证信息与收集到的所述用户信息之间的一致性指标,并保存得到的一致性计算结果。

[0144] 在本发明的一些实施例中,请参阅图4-d所示,所述众包项目信息处理模块402,包括:

[0145] 项目任务处理模块4021,用于确定所述众包项目的项目任务,获取所述项目任务的内容信息,所述内容信息包括:所述众包项目的名称、类型、知识领域范围、时间、费用、质量、资源、风险;

[0146] 项目信息分析模块4022,用于根据所述项目任务的内容信息估计所述项目任务所需的成本和所述项目任务的完成要求,其中,所述项目任务所需的成本包括:人力成本、资金成本和时间成本,所述人力成本包括:所述项目任务所需的专业技能类型、人员数量和专业技能水平,所述资金成本包括:所述项目任务的项目报酬、项目沟通渠道所需要的资金、项目资源提供成本、项目开销成本,所述时间成本包括:项目沟通成本和项目工时成本,所述完成要求包括:完成可靠度、完成质量、未完成和延期的损失。

[0147] 在本发明的一些实施例中,请参阅图4-e所示,所述众包参与者选取模块403,包括:

[0148] 用户意愿分析模块4031,用于根据所述每个众包参与人的信用信息和能力信息、所述项目任务的项目信息,预测所述每个众包参与人的项目接受可能性指标、项目中途放弃可能性指标;

[0149] 人力需求与参与者能力匹配模块4032,用于根据所述每个众包参与人的能力信息和所述每个众包参与人的项目接受可能性指标,从所述多个众包参与者中查找到符合所述项目任务所需的人力成本、且符合所述项目任务的完成质量的第一众包参与者子集;

[0150] 完成可靠度与参与者信用匹配模块4033,用于根据所述每个众包参与人的信用信息从所述多个众包参与者中查找到符合所述项目任务的完成可靠度的第二众包参与者子集;

[0151] 项目进度保证模块4034,用于根据所述每个众包参与人的能力信息和所述每个众包参与人的项目中途放弃可能性指标,从所述多个众包参与者中查找到符合所述项目任务所需的时间成本、资金成本、且符合所述项目任务的未完成和延期的损失的第三众包参与者子集;

[0152] 候选人集合确定模块4035,用于根据所述第一众包参与者子集、所述第二众包参与者子集、所述第三众包参与者子集确定出与所述项目信息匹配的众包参与者作为候选人集合。

[0153] 在本发明的一些实施例中,所述完成可靠度与参与者信用匹配模块4033,具体用于若所述一致性计算结果表示众包参与者的个人认证信息与收集到的用户信息存在不一致时,将所述一致性计算结果存储为反欺诈识别信息;根据所述每个众包参与者的信用信

息和所述反欺诈识别信息从所述多个众包参与人中查找到符合所述项目任务的完成可靠度的第二众包参与人子集。

[0154] 通过以上实施例对本发明实施例的描述可知,根据多个众包参与人分别提供的各自的行为记录和信用记录,获取多个众包参与人中每个众包参与人的信用信息和能力信息,从众包项目平台中确定需要分配参与人的众包项目,并从众包项目中获取到项目信息,项目信息包括:众包项目的项目任务以及项目任务的内容信息,最后可以根据每个众包参与人的信用信息和能力信息为众包项目选取与项目信息匹配的众包参与人作为候选人集合,候选人集合包括:从多个众包参与人中筛选出的用于执行项目任务的众包参与人。本发明实施例中可以根据众包参与人的行为记录和信用记录确定出该参与人的信用信息和能力信息,从而保证对众包参与人的真实刻画,保证众包参与人存在欺诈的可能性足够小,并可以准确地描述出参与人的能力,对于需要分配参与人的众包项目,能够对该众包项目的项目任务做出解析,从而确定出项目任务的内容信息,该内容信息可以反映出该项目任务对参与人的要求,最后再通过前述步骤获取到的众包参与人的信用信息和能力信息,与项目任务进行匹配,筛选出的候选人集合中的众包参与人可以执行该项目任务,从而实现为众包项目有效选取参与人,保证筛选的可靠性,这与现有技术中依靠金钱担保选取参与人的方式是完全不同的,本发明实施例可以提升众包项目平台中参与人选取的效率。

[0155] 图5是本发明实施例提供的一种服务器结构示意图,该服务器1100可因配置或性能不同而产生比较大的差异,可以包括一个或一个以上中央处理器(central processing units,CPU) 1122(例如,一个或一个以上处理器)和存储器1132,一个或一个以上存储应用程序1142或数据1144的存储介质1130(例如一个或一个以上海量存储设备)。其中,存储器1132和存储介质1130可以是短暂存储或持久存储。存储在存储介质1130的程序可以包括一个或一个以上模块(图示没标出),每个模块可以包括对服务器中的一系列指令操作。更进一步地,中央处理器1122可以设置为与存储介质1130通信,在服务器1100上执行存储介质1130中的一系列指令操作。

[0156] 服务器1100还可以包括一个或一个以上电源1126,一个或一个以上有线或无线网络接口1150,一个或一个以上输入输出接口1158,和/或,一个或一个以上操作系统1141,例如Windows Server™,Mac OS X™,Unix™,Linux™,FreeBSD™等等。

[0157] 上述实施例中由服务器所执行的众包项目中选取众包参与人的方法步骤可以基于该图5所示的服务器结构。

[0158] 另外需说明的是,以上所描述的装置实施例仅仅是示意性的,其中所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。另外,本发明提供的装置实施例附图中,模块之间的连接关系表示它们之间具有通信连接,具体可以实现为一条或多条通信总线或信号线。本领域普通技术人员在不付出创造性劳动的情况下,即可以理解并实施。

[0159] 通过以上的实施方式的描述,所属领域的技术人员可以清楚地了解到本发明可借助软件加必需的通用硬件的方式来实现,当然也可以通过专用硬件包括专用集成电路、专用CPU、专用存储器、专用元器件等来实现。一般情况下,凡由计算机程序完成的功能都可以

很容易地用相应的硬件来实现,而且,用来实现同一功能的具体硬件结构也可以是多种多样的,例如模拟电路、数字电路或专用电路等。但是,对本发明而言更多情况下软件程序实现是更佳的实施方式。基于这样的理解,本发明的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来,该计算机软件产品存储在可读取的存储介质中,如计算机的软盘、U盘、移动硬盘、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、磁碟或者光盘等,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备等)执行本发明各个实施例所述的方法。

[0160] 综上所述,以上实施例仅用以说明本发明的技术方案,而非对其限制;尽管参照上述实施例对本发明进行了详细的说明,本领域的普通技术人员应当理解:其依然可以对上述各实施例所记载的技术方案进行修改,或者对其中部分技术特征进行等同替换;而这些修改或者替换,并不使相应技术方案的本质脱离本发明各实施例技术方案的精神和范围。

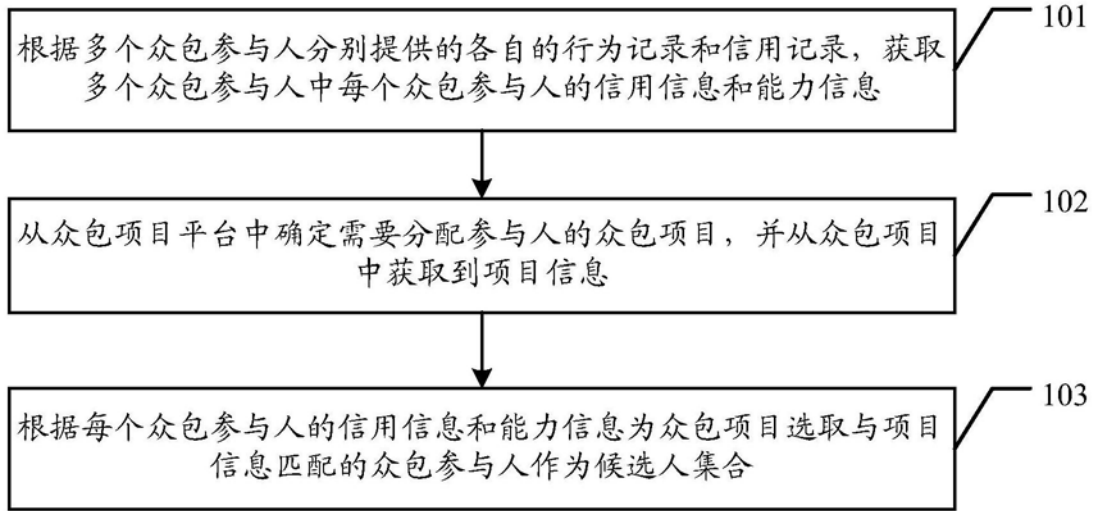


图1

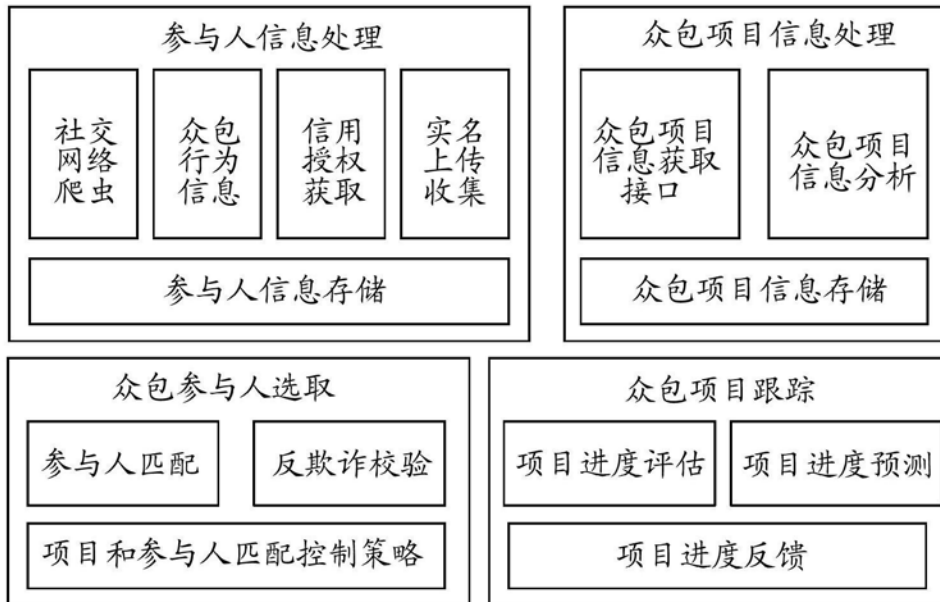


图2

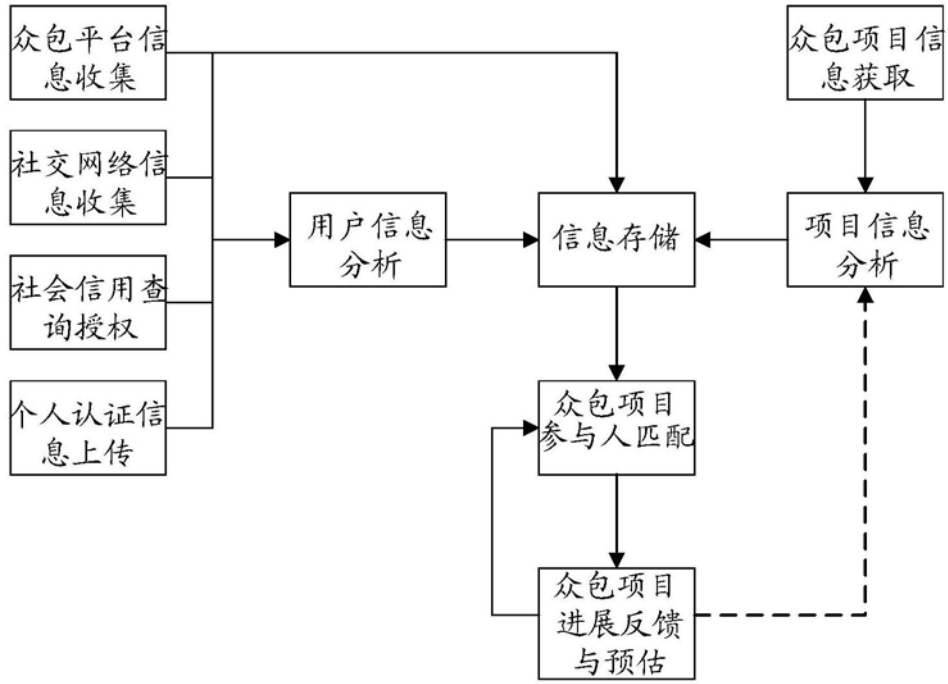


图3

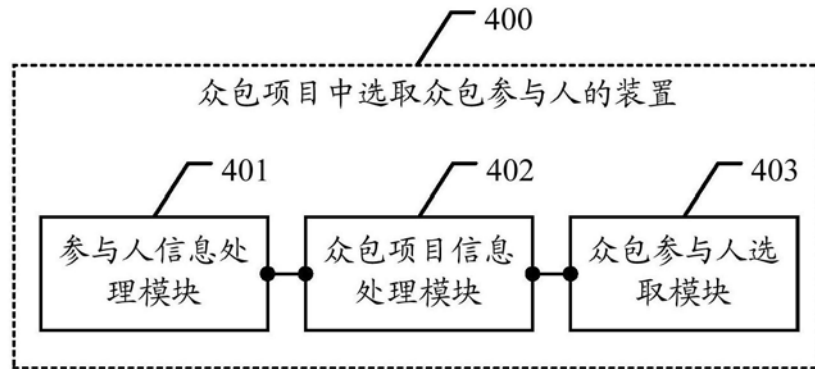


图4-a

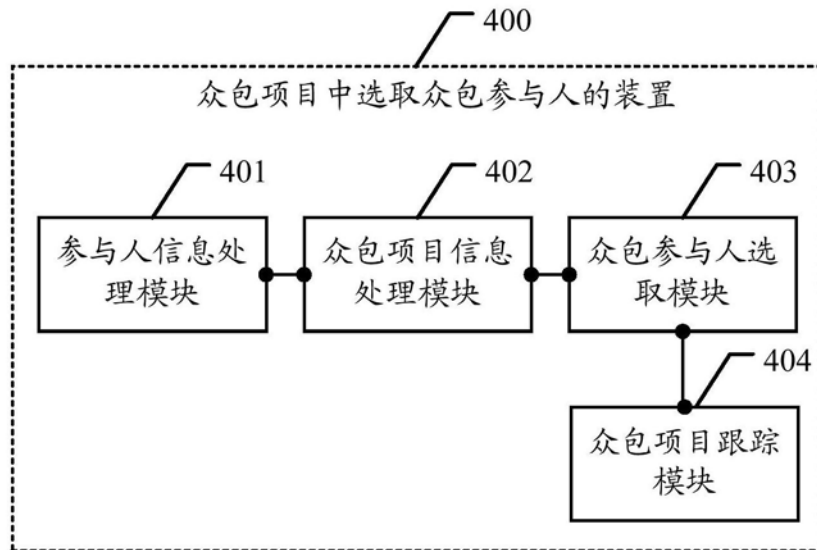


图4-b

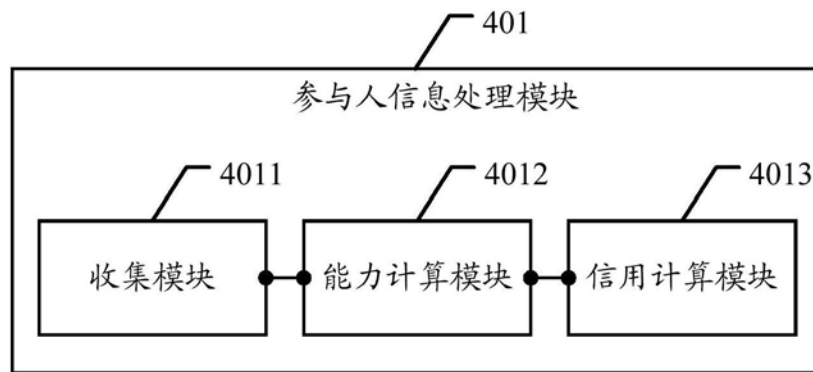


图4-c

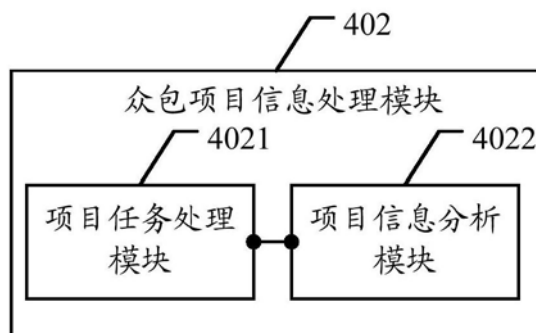


图4-d

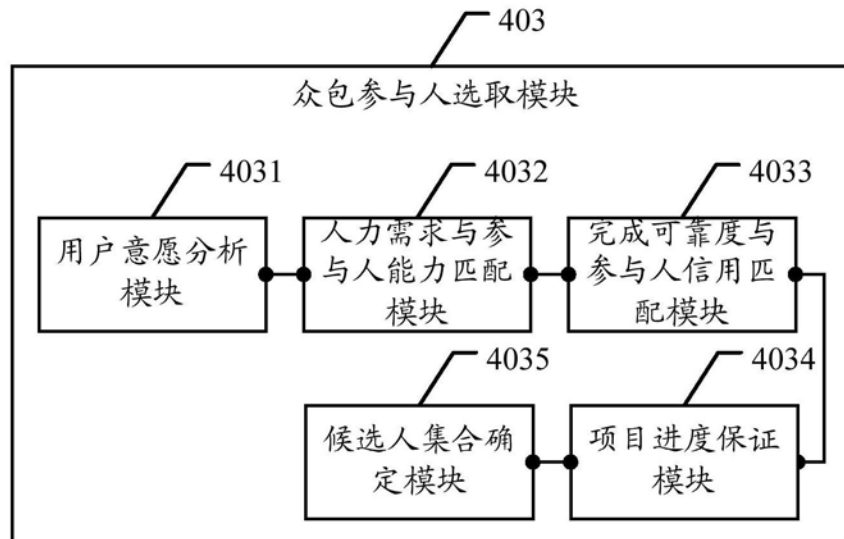


图4-e

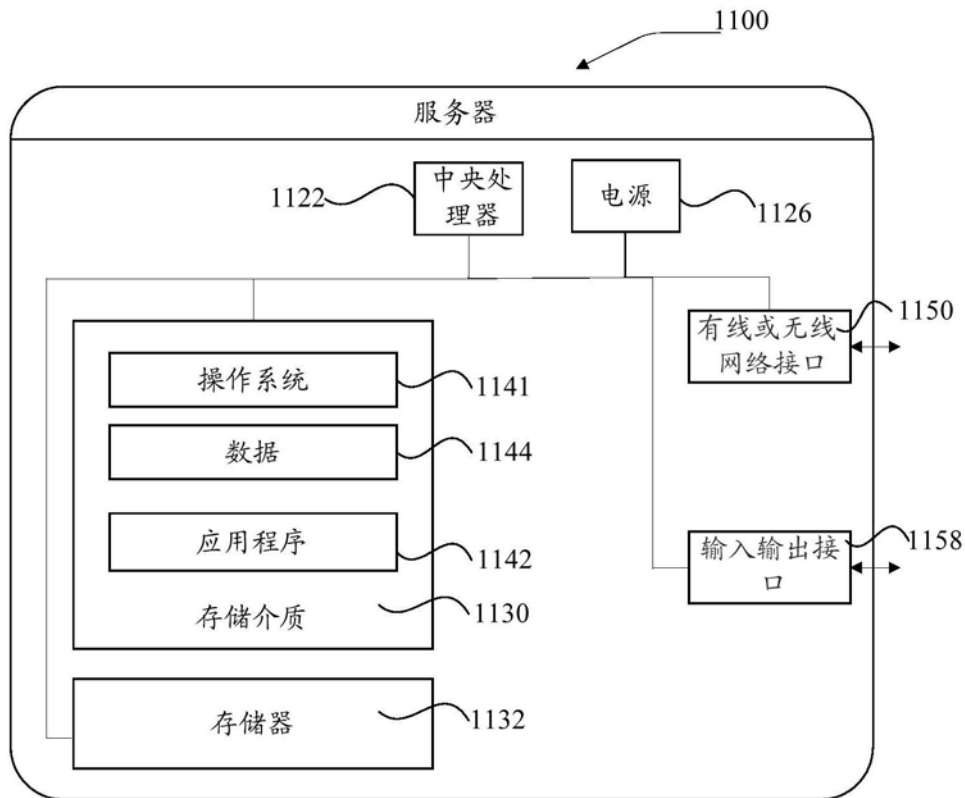


图5