



(12)发明专利申请

(10)申请公布号 CN 110705889 A

(43)申请公布日 2020.01.17

(21)申请号 201910958816.8

(22)申请日 2019.10.10

(71)申请人 华夏幸福产业投资有限公司
地址 100160 北京市丰台区南四环西路188号十八区25号楼3层301(园区)

(72)发明人 赵威

(74)专利代理机构 北京品源专利代理有限公司
11332

代理人 孟金喆

(51) Int. Cl.

G06Q 10/06(2012.01)

G06K 9/62(2006.01)

G06N 3/04(2006.01)

G06Q 50/26(2012.01)

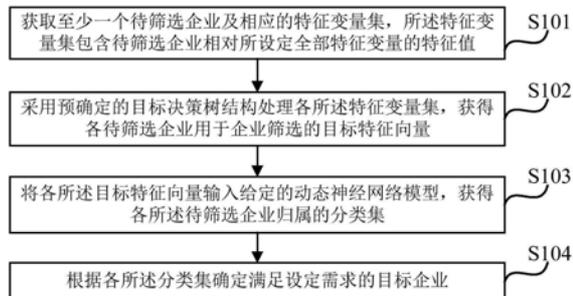
权利要求书3页 说明书12页 附图3页

(54)发明名称

一种企业筛选方法、装置、设备及存储介质

(57)摘要

本发明公开了一种企业筛选方法、装置、设备及存储介质,该方法包括:获取至少一个待筛选企业及相应的特征变量集,特征变量集包含待筛选企业相对所设定全部特征变量的特征值;采用预确定的目标决策树结构处理各特征变量集,获得各待筛选企业用于企业筛选的目标特征向量;将各所述目标特征向量输入给定的动态神经网络模型,获得各所述待筛选企业归属的分类集;根据各所述分类集确定满足设定需求的目标企业。利用该方法,规避了现有需要大量专家、时效低以及评估维度片面的问题,实现了企业的智能化筛选,且通过对企业特征变量集的处理所确定的目标特征向量的方式更好的规避了无效特征变量对企业筛选结果的影响,有效保证企业筛选的准确率。



1. 一种企业筛选方法,其特征在于,包括:

获取至少一个待筛选企业及相应的特征变量集,所述特征变量集包含待筛选企业相对所设定全部特征变量的特征值;

采用预确定的目标决策树结构处理各所述特征变量集,获得各待筛选企业用于企业筛选的目标特征向量;

将各所述目标特征向量输入给定的动态神经网络模型,获得各所述待筛选企业归属的分类集,所述动态神经网络模型采用所述目标决策树结构对应的特征变量进行模型训练;

根据各所述分类集确定满足设定需求的目标企业。

2. 根据权利要求1所述的方法,其特征在于,所述目标决策树结构的确定过程包括下述步骤:

获取包含至少一个样本企业的样本企业集,初始化参与决策的特征变量的决策数量,各所述样本企业对应存在标准特征变量集及标准所属分类;

从所设定全部特征变量中选定包含所述决策数量个特征变量的决策特征集;

基于所述样本企业集训练形成所述决策特征集对应的决策树结构,并获得各样本企业在所述决策树结构下的目标所属分类;

根据各样本企业的标准所属分类及在所述决策树结构下对应的目标所属分类,确定所述决策树结构的分类准确率;

若所述分类准确率达到设定阈值,则将所述决策树结构确定为目标决策树结构;否则,将预设定的选定次数加1,返回执行决策特征集的选定操作;

直至所述选定次数达到设定次数阈值后,更新所述决策数量并返回决策特征集的选定操作。

3. 根据权利要求2所述的方法,其特征在于,所述所设定全部特征变量按照特征价值优先级分布在至少一个特征分类集;

相应的,所述从所设定全部特征变量中选定包含所述决策数量个特征变量的决策特征集,包括:

根据所述决策数量及所述特征分类集的个数,确定每个特征分类集中特征变量的选定量;

从各所述特征分类集中选取一组由所述选定量个特征变量构成的未选定过的特征组合,基于各所述特征组合形成决策特征集。

4. 根据权利要求3所述的方法,其特征在于,所述决策数量以步长为1更新;

相应的,所述根据所述决策数量及所述特征分类集的个数,确定每个特征分类集中特征变量的选定量,包括:

当所述决策数量整除所述特征分类集的个数时,将整除的商作为每个特征分类集中特征变量的选定量;否则,

将所述决策数量与所述特征分类集的商作为每个特征分类集中特征变量当前的选定量;

将按优先级由高到低排列的设定个数特征分类集对应的选定量加1,其中,所述设定个数为所述决策数量与所述特征分类集的余数。

5. 根据权利要求2所述的方法,其特征在于,所述基于所述样本企业集训练形成所述决

策特征集对应的决策树结构,并获得各样本企业在所述决策树结构下的目标所属分类,包括:

以待构建树结构第一层的根节点为当前节点,并将所述决策特征集作为当前的候选特征集;

针对所述候选特征集中的每个决策特征变量,获取所述样本企业集中各样本企业对应所述决策特征变量的决策特征值;

基于各所述决策特征值结合给定的信息增益计算公式,确定所述决策特征变量相对所述当前节点的信息增益率;

比对各所述决策特征变量相对所述当前节点的信息增益率,并将最高信息增益率对应的决策特征变量作为所述当前节点的关键特征变量;

根据所述关键特征变量从所述决策特征集中确定待构建树结构下一层对应候选特征变量集以及下一层所包含节点的节点数量;

将所述节点数量的子节点分别作为新的当前节点,返回执行决策特征值的获取操作直至所述候选特征变量集为空;

获得构建的待剪枝决策树结构,并采用各所述样本企业对所述待剪枝决策树结构进行剪枝处理形成精简的决策树结构;

根据所述决策树结构中各叶节点对应的数据信息,各样本企业在所述决策树结构下的目标所属分类。

6. 根据权利要求1-5任一项所述的方法,其特征在于,所述采用预确定的目标决策树结构处理各所述特征变量集,获得各待筛选企业用于企业筛选的目标特征向量,包括:

将各所述特征变量集输入具备所述目标决策树结构的决策树模型;

获得从各所述特征变量集筛选的以所述目标决策树结构排序的目标特征变量序列,各所述目标特征变量序列与所述待筛选企业一一对应;

基于各所述目标特征变量序列中对应的特征值,构成各所述待筛选企业用于企业筛选的目标特征向量。

7. 根据权利要求1-5任一项所述的方法,其特征在于,在将各所述目标特征向量输入给定的动态神经网络模型,获得各所述待筛选企业归属的分类集之前,还包括:

向量化处理各所述待筛选企业对应的企业评价文本,获得各所述企业评价文本的企业评价向量;

对各所述待筛选企业的目标特征向量及企业评价向量进行合并处理,将合并后形成的向量记为新的目标特征向量。

8. 根据权利要求1-5任一项所述的方法,其特征在于,还包括:

处理所抓取的至少一个企业的原始数据,从处理后的数据中提取所设定全部特征变量的特征值,形成各所述企业的特征变量集。

9. 一种企业筛选装置,其特征在于,包括:

获取模块,用于获取至少一个待筛选企业及相应的特征变量集,所述特征变量集包含待筛选企业相对所设定全部特征变量的特征值;

特征处理模块,用于采用预确定的目标决策树结构处理各所述特征变量集,获得各待筛选企业用于企业筛选的目标特征向量;

企业分类模块,用于将各所述目标特征向量输入给定的动态神经网络模型,获得各所述待筛选企业归属的分类集;

目标企业确定模块,用于根据各所述分类集确定满足设定需求的目标企业。

10. 一种企业筛选设备,其特征在于,包括:存储器以及一个或多个处理器;

所述存储器,用于存储一个或多个程序;

当所述一个或多个程序被所述一个或多个处理器执行,使得所述一个或多个处理器实现如权利要求1-8中任一所述的企业筛选方法。

11. 一种包含计算机可执行指令的存储介质,其特征在于,所述计算机可执行指令在由计算机处理器执行时用于执行如权利要求1-8中任一所述的企业筛选方法。

一种企业筛选方法、装置、设备及存储介质

技术领域

[0001] 本发明实施例涉及数据处理技术领域,尤其涉及一种企业筛选方法、装置、设备及存储介质。

背景技术

[0002] 为了加快地区的经济发展,通过招商引资的方式来引进企业已成为重要一环,其重要性日益突出。为提高招商企业的有效性,在企业招商过程中,往往通过对企业招商价值的评定来选择高价值的企业。

[0003] 目前,企业招商价值的评定往往通过人工筛选的方式来进行,如由专家基于专家来对有入驻意向的企业进行筛选,然而,随着招商业务多样性和复杂度的不断升级,专家的经验就显得比较单薄,除此之外,企业选择的时效性和全面性也急需提升。

发明内容

[0004] 本发明实施例提供一种企业筛选方法、装置、设备及存储介质,实现了简单、快速、精准的预测获得满足招商需求的高招商价值的企业。

[0005] 第一方面,本发明实施例提供了一种企业筛选方法,包括:

[0006] 获取至少一个待筛选企业及相应的特征变量集,所述特征变量集包含待筛选企业相对所设定全部特征变量的特征值;

[0007] 采用预确定的目标决策树结构处理各所述特征变量集,获得各待筛选企业用于企业筛选的目标特征向量;

[0008] 将各所述目标特征向量输入给定的动态神经网络模型,获得各所述待筛选企业归属的分类集;

[0009] 根据各所述分类集确定满足设定需求的目标企业。

[0010] 第二方面,本发明实施例还提供了一种企业筛选装置,包括:

[0011] 获取模块,用于获取至少一个待筛选企业及相应的特征变量集,所述特征变量集包含待筛选企业相对所设定全部特征变量的特征值;

[0012] 特征处理模块,用于采用预确定的目标决策树结构处理各所述特征变量集,获得各待筛选企业用于企业筛选的目标特征向量;

[0013] 企业分类模块,用于将各所述目标特征向量输入给定的动态神经网络模型,获得各所述待筛选企业归属的分类集;

[0014] 目标企业确定模块,用于根据各所述分类集确定满足设定需求的目标企业。

[0015] 第三方面,本发明实施例还提供了一种企业筛选设备,包括:存储器以及一个或多个处理器;

[0016] 所述存储器,用于存储一个或多个程序;

[0017] 当所述一个或多个程序被所述一个或多个处理器执行,使得所述一个或多个处理器实现如第一方面所述的企业筛选方法。

[0018] 第四方面,本发明实施例还提供了一种包含计算机可执行指令的存储介质,所述计算机可执行指令在由计算机处理器执行时用于执行如第一方面所述的企业筛选方法。

[0019] 本发明实施例提供了一种企业筛选方法、装置、设备以及存储介质,该方法包括:获取至少一个待筛选企业及相应的特征变量集,所述特征变量集包含待筛选企业相对所设定全部特征变量的特征值;采用预确定的目标决策树结构处理各所述特征变量集,获得各待筛选企业用于企业筛选的目标特征向量;将各所述目标特征向量输入给定的动态神经网络模型,获得各所述待筛选企业归属的分类集;根据各所述分类集确定满足设定需求的目标企业。上述技术方案,与现有通过专家经验筛选相比,规避了需要大量专家、时效低以及评估维度片面等问题,实现了招商企业的智能化筛选,且在智能化筛选中具体通过对待筛选企业所具备全部特征变量的筛选处理确定了有效的特征变量,来作为企业筛选所输入的目标特征向量,从而避免了无效特征变量的干扰,有效保证了企业筛选的准确率。

附图说明

- [0020] 图1是本发明实施例一提供的一种企业筛选方法的流程图;
- [0021] 图2是本发明实施例一提供的确定目标特征变量的流程示意图;
- [0022] 图3是本发明实施例二提供的一种企业筛选方法的流程图;
- [0023] 图4是本发明实施例三提供的一种企业筛选装置的结构框图;
- [0024] 图5是本发明实施例四提供的一种企业筛选设备的结构示意图。

具体实施方式

[0025] 下面结合附图和实施例对本发明作进一步的详细说明。可以理解的是,此处所描述的具体实施例仅仅用于解释本发明,而非对本发明的限定。另外还需要说明的是,为了便于描述,附图中仅示出了与本发明相关的部分而非全部结构。

[0026] 实施例一

[0027] 图1为本发明实施例一提供的一种企业筛选方法的流程示意图,该方法适用于对参与招商的企业进行筛选的情况,该方法可以有企业筛选装置来执行,该装置可以采用软件和/或硬件的方式实现,该装置可以配置于设备中,例如典型的是计算机等。如图1所示,该方法具体包括如下步骤:

[0028] S101、获取至少一个待筛选企业及相应的特征变量集,所述特征变量集包含待筛选企业相对所设定全部特征变量的特征值。

[0029] 需要说明的是,本实施例所提供企业筛选方法的应用场景可以是任何需要进行企业招商的应用领域,所述应用领域可以是互联网领域以及房地产领域等;所述待筛选企业具体可理解为在企业招商中有入驻意向的企业。示例性的,本步骤可直接从提供有企业入驻申请入口的平台获取至少一个待筛选企业或者从企业信息库中获取至少一个待筛选企业,其中,所述企业信息库具体可理解为存储了相关人员通过线上或线下收集的有入驻意向的企业。

[0030] 本步骤可直接获取到有入驻意向的待筛选企业,其中,所获取的待筛选企业可以通过企业名称或企业标识来表示,同时,本步骤还可获取各待筛选企业相应的预先经过处理后形成的特征变量集,所述特征变量集具体可理解为包含了待筛选企业相对所设定全部

特征变量的特征值,所述特征变量集具体可以是预先对所获取各待筛选企业的原始数据进行筛选、分类以及提取等处理操作后形成的集合。

[0031] 可以知道的是,在确定一个需要进行企业招商的应用领域后,根据该应用领域的特点和性质,与该应用领域相关的企业应该具备的特征变量也可对应确定下来,所确定的特征变量即可认为是本实施例上述提及特征向量集中的全部特征变量。示例性的,以房地产领域为例,在对新的商业楼盘进行企业引入时,所考虑的特征变量可以有基础变量、累计变量、异动变量以及序列变量等,其中,基础变量可以理解为对企业进行基础信息概述的特征变量,可以包括:企业名称、企业创始时间以及企业创始地点等;所述累计变量可以是企业生产价值的累计情况进行概述的特征变量,如企业运营的年度累计产值等;所述异动变量可理解为对企业的运营带来异动影响的特征变量,如消费者投诉等特征变量,所述序列变量具体可理解为对企业的发展及事件沿时间线描述的特征变量。

[0032] 可以理解的是,不同的企业根据自身发展的不同,所述待筛选企业所具备特征变量集中所包括所设定全部特征变量相对自身的发展具备与其他企业不同的特征值,各所述待筛选企业的特征变量集可作为企业筛选的输入信息。

[0033] S102、采用预确定的目标决策树结构处理各所述特征变量集,获得各待筛选企业用于企业筛选的目标特征向量。

[0034] 需要说明的是,本实施例在通过特征变量集进行企业筛选前,首先需要基于本步骤对各待筛选企业的特征变量集进行筛选处理,以从所述设定全部变量集中筛选出对企业的筛选结果影响力度较高的目标特征变量,由此可获得各待筛选企业相对各目标特征变量的特征值,并获得通过各待筛选企业相对各目标特征变量的特征值构成的目标特征向量,所述目标特征向量具体可理解为用于作为企业筛选所需的输入信息。

[0035] 在本实施例中,所述目标决策树结构具体可理解为对特征变量集中包括的部分特征变量经过树结构的构造策略处理后形成的树形结构,该目标决策树结构中所有的非叶节点可理解为选定的特征变量,每个各非叶节点对应的特征变量具有排列顺序限定,此外,该目标决策树结构中所有的叶节点可理解为基于上述排列顺序的特征变量对一个待分类集合进行分类处理形成的分类结果,本步骤可以将待筛选企业及特征变量集经过目标决策树结构的处理,确定出作为该目标决策树结构中各非叶节点的目标特征变量,同时,还可以获得的是根据所确定出中的各目标特征变量对至少一个待筛选企业进行分类处理后得到的分类结果。经过本步骤的处理后,可以确定待筛选企业相对各目标特征变量的特征值,基于各目标特征变量的特征值,相对各待筛选企业可以形成以所具备目标特征变量的个数作为向量维度的目标特征向量。

[0036] 在本实施例中,所述目标决策树结构的建立过程可以通过给定的样本企业集对决策树模型的训练学习确定。进一步地,图2给出了本发明实施例一中确定目标决策树结构的流程示意图,如图2所示,目标决策树结构的确定过程具体包括如下步骤:

[0037] S1001、获取包括至少一个样本企业的样本企业集,初始化参与决策的特征变量的决策数量,其中,各所述样本企业对存在标准特征变量集及标准所属分类。

[0038] 在本实施例中,所述样本企业集包括至少一个样本企业,所述样本企业可以是根据历史企业招商的招商结果收集到的处于不同分类层次的历史企业,各所述样本企业都相应存在一个标准特征变量集以及标准所属分类,所述标准特征变量集具体可理解为包含了

所设定全部特征变量所对应标准特征值的集合,所述标准所属分类具体可理解为在历史企业招商中确定出的分类结果。

[0039] 本实施例设定一个决策数量作为目标决策树结构确定中的一个变量,所述决策数量具体用于限定目标决策树结构确定时每次参与决策树构建的特征变量的数量。

[0040] S1002、从所设定全部特征变量中选定包含所述决策数量个特征变量的决策特征集。

[0041] 在本实施例中,所述决策特征集具体可理解为当前参与决策树构建的所有特征变量的集合,所述决策特征集所包含特征变量的数量为所述决策数量个,具体通过本步骤从所设定全部特征变量中选定。示例性的,本步骤选定特征变量的方式可以是顺序依次选取,也可以是按照全部特征变量的特点以及所具备的价值采用一定的选取策略选取。

[0042] 可选的,所述所设定全部特征变量按照特征价值优先级分布在至少一个特征分类集,相应的,本实施例具体将从所设定全部特征变量中选定包含所述决策数量个特征变量的决策特征集,优化为:根据所述决策数量及所述特征分类集的个数,确定每个特征分类集中特征变量的选定量;从各所述特征分类集中选取一组由所述选定量个特征变量构成的未选定过的特征组合,基于各所述特征组合形成决策特征集。

[0043] 具体的,本实施例考虑首先将所设定全部特征变量按照特征价值优先级进行特征分类,其中,每个特征变量的优先级可以结合该特征变量在历史企业招商中进行企业分类时的贡献确定,具体的,可以在每次企业筛选完成后,对所采用的的特征变量进行特征价值计算,所述特征价值具体可理解为特征变量在企业筛选过程中的权重值,特征变量的价值越高可认为所述特征变量的优先级越高,之后可根据优先级的不同,将全部特征变量划分到相应的特征分类集中。

[0044] 在本实施例中,所述特征分类集的个数与预确定的特征价值优先级的级数对应,将全部特征变量划分到不同的特征分类集后,常规的特征变量的选定操作为直接从特征价值高优先级对应的特征分类集中选定,本实施例考虑到特征变量对于不同待筛选企业的影响力度不同,采用从各所述特征分类集中分别选取特征变量的方式来均衡特征变量对待筛选企业的影响。具体的,本实施例首先基于决策数量和特征分类集的个数来确定需要从每个特征分类集中选取特征变量的选定量,然后选取出选定量的特征变量来构成一组特征组合,最终将各特征分类集对应的特征组合再次添加合并构成决策特征集。

[0045] 需要说明的是,本实施例从各特征分类集中选取特征变量时需要满足当前选取形成的特征组合时之前未选定过的,示例性的,假设一个特征变量集中包含了A、B、C、D、E以及F,若上一次选取形成的特征组合为{A、B、C},则后续选取形成的组合中将不再选定该特征组合,所选的特征组合可以是{A、B、D}或者{B、C、D}等存在重复选取的特征变量,但最终形成的特征组合是之前未选定过的。本实施例可以采用特征变量优先级由高到低的顺序对多个特征组合进行合并。

[0046] 在上述优化的基础上,所述决策数量以步长为1更新,相应的,本实施例将根据所述决策数量及所述特征分类集的个数,确定每个特征分类集中特征变量的选定量的过程具体优化为:当所述决策数量整除所述特征分类集的个数时,将整除的商作为每个特征分类集中特征变量的选定量;否则,将所述决策数量与所述特征分类集的商作为每个特征分类集中特征变量当前的选定量;将按优先级由高到低排列的设定个数特征分类集对应的选定

量加1,其中,所述设定个数为所述决策数量与所述特征分类集的余数。

[0047] 在本实施例中,所述决策变量为目标决策树结构确定过程中的一个循环变量,在满足决策变量的更新条件时,决策变量将以步长为1进行更新。确定每个特征分类集所对应选定量时,考虑每个特征分类集对应的选定量基本相同,由此,首先考虑决策数量是否能够基于特征分类集的个数进行等分,若能,则等分从每个特征分类集中选取的特征变量的数量,示例性的,当特征分类集的个数为3,决策数量为9时,可以确定每个特征分类集对应的选定量为3;因为决策数量在发生变化,因此存在决策数量与上述个数如果无法等分的情况,则此时可考虑先按优先级由高到低选定余数个特征分类集,并将选定的各特征分类集的选定量在当前基础上加1。

[0048] 由此,上述方式既考虑了特征变量在各特征分类集中的均衡分配,同时也保证了高优先级的特征变量的优先选。同时,可以理解的是,基于上述操作确定选定量时,若确定的选定量大于特征分类集本身所包含特征变量的个数,则该特征分类集的选定量将等于所包含特征变量的个数,并将一个选定特征分类集对应的选定量的加1,由此循环类推,保证总的决策数量统一,其中,选定特征分类集的选定规则可以是选定与上述超过所包含特征变量个数的特征分类集相邻排列的下一个特征分类集,且若上述超过所包含特征变量个数的特征分类集排在末位,则将首个特征分类集作为选定特征分类集。

[0049] S1003、基于所述样本企业集训练形成所述决策特征集对应的决策树结构,并获得各样本企业在所述决策树结构下的目标所属分类。

[0050] 在本实施例中,基于上述步骤确定决策特征集后,最关键的步骤就是基于决策特征集中的特征变量通过构造决策树的方式来进行各特征变量在树结构上的排序,即如何将每个特征变量赋予到决策树中的节点上。具体的,本步骤可从样本企业集所包括每个样本企业对应的标准特征变量集中筛选属于决策特征集的特征变量及其相应的标准特征值,并基于各标准特征值构成标准特征向量,将具备标准特征向量的样本企业作为训练数据,然后基于决策树的构造规则,确定决策特征集中各特征变量在决策树各节点上的分布,由此形成基于决策树中各特征变量形成的决策树结构。

[0051] 同时,可以理解的是,在基于决策变量集构造决策树的过程中也相应对样本企业按照所对应的标准特征向量进行了分类,由此在决策树构造完成时,可以相应获得各样本企业在该决策树结构在分类处理后形成的目标所属分类。

[0052] 进一步地,本实施例将基于所述样本企业集训练形成所述决策特征集对应的决策树结构,并获得各样本企业在所述决策树结构下的目标所属分类,具体优化为:

[0053] a) 以待构建树结构第一层的根节点为当前节点,并将所述决策特征集作为当前的候选特征集。

[0054] 本步骤相当于一个初始化步骤,以用于启动决策树结构的构建。

[0055] b) 针对所述候选特征集中的每个决策特征变量,获取所述样本企业集中各样本企业对应所述决策特征变量的决策特征值。

[0056] c) 基于各所述决策特征值结合给定的信息增益计算公式,确定所述决策特征变量相对所述当前节点的信息增益率。

[0057] d) 比对各所述决策特征变量相对所述当前节点的信息增益率,并将最高信息增益率对应的决策特征变量作为所述当前节点的关键特征变量。

[0058] 在本实施例中,所述信息增益率具体可用来体现当前的决策特征变量处于当前节点下对整个分类结果的影响程度,信息增益率越高,表明该信息增益率处于当前节点的可能性越大。

[0059] e) 根据所述关键特征变量从所述决策特征集中确定待构建树结构下一层对应候选特征变量集以及下一层所包含节点的节点数量。

[0060] 在本实施例中,处于所述当前节点的关键特征变量会对下一层节点的选定产生影响,即,在当前节点以某个决策特征变量作为关键特征变量时,可形成的下一层节点的数量可以确定,且能够作为下一层节点的决策特征变量的范围也相应缩小。示例性的,可以按照设定的阈值对现有的关键特征变量的进行分裂,从而形成多个可作为下一层的分支,并将分支数量作为下一层的节点数量,同时可也根据关键特征变量与决策特征集中其他特征变量的关联性来确定下一层对应的候选特征变量集。

[0061] f) 将所述节点数量的子节点分别作为新的当前节点,返回执行b) 直至当前层的各节点具备同一属性。

[0062] g) 获得构建的待剪枝决策树结构,并采用各所述样本企业对所述待剪枝决策树结构进行剪枝处理形成精简的决策树结构。

[0063] 在本实施例中,所述剪枝处理具体可理解为对上述所形成的决策树结构的精简处理。

[0064] h) 根据所述决策树结构中各叶节点对应的数据信息,各样本企业在所述决策树结构下的目标所属分类。

[0065] S1004、根据各样本企业的标准所属分类及在所述决策树结构下对应的目标所属分类,确定所述决策树结构的分类准确率。

[0066] 在本实施例中,每个样本企业对应存在标准所属分类,基于上述S1003还可确定出所述样本企业在当前所构造决策树结构下的目标所属分类,本步骤可将标准属实分类与所述目标所属分类进行结果匹配,从而确定出决策树结构的分类准确率。

[0067] S1005、确定所述分类准确率是否达到设定阈值,若是,则执行S1006;若否,则执行S1007。

[0068] 需要说明的是,现有的决策树训练确定最终决策树结构的方法中,往往在最初给定循环迭代次数后,会一直进行迭代训练直至达到循环迭代次数,在每次迭代完成后同样会进行获得一个分类准确率,由此在循环迭代结束后,通过比对各分类准确率的高低来算出训练处的目标决策树结构。

[0069] 本实施例中,考虑在完成一次迭代训练后将相应的分类准确率直接与设定阈值进行比对,如果没有达到设定阈值,则可执行S1007继续进行目标决策树结构的确定,若达到设定阈值,则可执行S1006将当前形成的决策树结构作为目标决策树结构,由此在后续基于样本数据进行决策树模型训练中不再需要重新确定决策树结构,可以直接采用确定的目标决策树结构来形成样本向量,示例性的,所述设定阈值可以是一个为80%的比率值。

[0070] S1006、将所述决策树结构确定为目标决策树结构。

[0071] 需要说明的是,本实施例提供的目标决策树结构的确定方式,降低了目标决策树结构确定过程中的计算复杂度,提高了计算时效,同时在实际操作中,采用上述方式确定目标决策树结构的计算耗费时间可由原来的一周缩减到当前的1小时,很大程度减少了资源

消耗。本步骤为上述S1005判断中的其中一个执行分支,当确定分类准确率达到了设定阈值时,可以将当前确定的决策树结构作为目标决策数结构。

[0072] S1007、判定预设定的选定次数是否达到设定次数阈值,若否,则执行S1008;若是,则执行S1009。

[0073] 在本实施例中,当分类准确率为达到设定阈值时,则可进一步执行决策树确定的循环操作,具体的,本步骤首先可判定预先给定的选定次数是否达到设定次数阈值,所述选定次数具体可理解为上述进行决策特征集选定的次数,最初可设置为0。所述设定次数阈值可以是一个经验值,也可以是基于组合公式确定的选定总次数。

[0074] 可选的,本实施例采用组合公式来确定设定次数阈值,具体的,所述设定次数阈值的确定与各特征分类集中所包括特征变量的个数以及所对应的选定量有关,示例性的,假设特征分类集A、B以及C包含的特征变量的个数分别为i、j和k,对应的选定量均为a,则所述设定次数阈值等于 $C_i^a \times C_j^a \times C_k^a$ 。此外,可以理解的是,本实施例中所述设定次数阈值随着决策数量及各特征分类集所对应选定量的更新相应更新。

[0075] S1008、将预设定的选定次数加1,返回执行S1002。

[0076] S1009、更新所述决策数量并返回执行S1002。

[0077] 示例性的,本实施例优选以步长为1进行决策数量的更新。

[0078] S103、将各所述目标特征向量输入给定的动态神经网络模型,获得各所述待筛选企业归属的分类集。

[0079] 在本实施例中,上述S102相当于对待筛选企业所关联特征数据的处理,从而获得作为输入数据的目标特征向量,所述动态神经网络模型具体可理解为预先训练的用于企业分类的机器学习模型,其训练样本可以是历史企业招商中已经具备分类结果的企业,具体的训练数据与上述进行目标决策树结构确定时所采用的可以是同一个。需要说明的是,为保证动态神经网络训练的准确度,本实施例进行所述动态神经网络模型训练学习是,优选采用所述目标决策树结构对应的特征变量进行模型训练。

[0080] 可只知道的是,动态神经网络模型的输出主要通过对待筛选企业招商价值评测后的分类输出,具体可包括强、中、弱三个分类集,由此可将招商价值高的企业划分到强分类集中。

[0081] S104、根据各所述分类集确定满足设定需求的目标企业。

[0082] 在本实施例中,所述设定需求可认为是招商者的招商意向需求,具体可以根据企业筛选应用领域的不同进行不同的设定,站在商业价值角度考虑,本实施例可以将筛选后划分到强分类集中的选企业作为目标企业。

[0083] 本发明实施例一提供的一种企业筛选方法,与现有通过专家经验筛选相比,规避了需要大量专家、时效低以及评估维度片面等问题,实现了招商企业的智能化筛选,且在智能化筛选中具体通过对待筛选企业所具备全部特征变量的筛选处理确定了有效的特征变量,来作为企业筛选所输入的目标特征向量,从而避免了无效特征变量的干扰,有效保证了企业筛选的准确率。

[0084] 实施例二

[0085] 图3为本发明实施例二提供的一种企业筛选方法的流程示意图,本实施例二在上

述实施例的基础上进行优化,在本实施例中,将采用预确定的目标决策树结构处理各所述特征变量集,获得各待筛选企业用于企业筛选的目标特征向量,进一步优化为:将各所述特征变量集输入具备所述目标决策树结构的决策树模型;获得从各所述特征变量集筛选的以所述目标决策树结构排序的目标特征变量序列,各所述目标特征变量序列与所述待筛选企业一一对应;基于各所述目标特征变量序列中对应的特征值,构成各所述待筛选企业用于企业筛选的目标特征向量。

[0086] 同时,本实施例在将各所述目标特征向量输入给定的动态神经网络模型,获得各所述待筛选企业归属的分类集之前,还优化包括:向量化处理各所述待筛选企业对应的企业评价文本,获得各所述企业评价文本的企业评价向量;对各所述待筛选企业的目标特征向量及企业评价向量进行合并处理,将合并后形成的向量记为新的目标特征向量。

[0087] 此外,本实施例还优化包括了:处理所抓取的至少一个企业的原始数据,从处理后的数据中提取所设定全部特征变量的特征值,形成各所述待筛选企业的特征变量集。

[0088] 如图3所示,本实施例二提供的一种企业筛选方法,具体包括如下操作:

[0089] S200、处理所抓取的至少一个企业的原始数据,从处理后的数据中提取所设定全部特征变量的特征值,形成各所述企业的特征变量集。

[0090] 本步骤相当于本实施例的一个预处理步骤,所述原始数据可理解为进行入驻申请时提供或填报的企业相关数据,本步骤可以通过数据抓取工具来获取各待筛选企业的原始数据,并对获取的各原始数据进行清洗,如检查数据一致性,处理无效值和缺失值等,然后从清洗后的数据中提取各特征变量的特征值,以次形成相应待筛选企业的特征变量集。

[0091] S201、获取各待筛选企业及相应的特征变量集,所述特征变量集包含待筛选企业相对所设定全部特征变量的特征值。

[0092] 示例性的,在上述步骤的基础上,本步骤可以从上述形成的企业集中选取至少一个待筛选企业及相应的特征变量集。

[0093] S202、将各所述特征变量集输入具备所述目标决策树结构的决策树模型。

[0094] 在本实施例中,S202至S204给出了目标特征向量的确定过程,本步骤中,所述决策树模型可理解为基于上述目标决策树结构的确定方式形成的包含了目标决策树结构的分类模型。本步骤中将所述特征变量集输入所述决策树模型的过程就可看做从特征变量集中筛选目标决策树结构所包含各目标特征向量的过程。

[0095] S203、获得从各所述特征变量集筛选的以所述目标决策树结构排序的目标特征变量序列,各所述目标特征变量序列与所述待筛选企业一一对应。

[0096] 在本实施例中,将各特征变量集输入上述决策模型后,可以保证各特征变量集中仅有属于目标决策树结构中的各目标特征变量参与企业筛选,通过上述S202的输入,可以获得各待筛选企业一一对应的目标特征变量序列。

[0097] S204、基于各所述目标特征变量序列中对应的特征值,构成各所述待筛选企业用于企业筛选的目标特征向量。

[0098] 在本实施例中,目标特征变量序列中包含了目标决策树结构中各目标特征变量,由此可从待筛选企业所对应各特征变量集中获得相应各目标特征变量的特征值,以形成用于企业筛选输入的目标特征向量。

[0099] S205、向量化处理各所述待筛选企业对应的企业评价文本,获得各所述企业评价

文本的企业评价向量。

[0100] 在本实施例中,上述目标特征向量仅包括了从待筛选企业的原始数据中提取出来的目标特征变量,具体由各目标特征变量的特征值构成,为提高企业筛选的准确率,本实施例进一步加入了企业评价文本形成的企业评价向量。其中,所述企业评价文本具体可理解为对待筛选企业进行评价所形成的包含企业评价关键词的文本信息,其对企业的评价可以由相关经验的专家通过人工标注的形式来完成,本步骤可直接获取到各待筛选企业对应的企业评价文本。

[0101] 本步骤可以对各企业评价文本进行向量化处理,其向量化的过程可描述为对企业评价文本进行分词处理,然后统计分词处理后各分词的出现频次,将频次大于设定阈值的分词作为企业评价的关键词,基于各所述关键词形成企业评价向量。

[0102] S206、对各所述待筛选企业的目标特征向量及企业评价向量进行合并处理,将合并后形成的向量记为新的目标特征向量。

[0103] 在本实施例中,可以将上述所确定目标特征向量与企业评价向量进行合并,从而作为新的输入动态神经网络模型的目标特征向量。需要说明的是,本实施例上述通过决策树模型输出的目标特征向量可以作为第一输入数据输入动态神经网络模型,在基于动态神经网络模型进行企业筛选处理之前,可以对第一输入数据进行维度扩展,并将企业评价向量添加进第一输入数据作为动态神经网络模型的输入。

[0104] S207、将各所述目标特征向量输入给定的动态神经网络模型,获得各所述待筛选企业归属的分类集。

[0105] S208、根据各所述分类集确定满足设定需求的目标企业。

[0106] 本发明实施例二提供的一种企业筛选方法,具体增加了待筛选企业原始数据的处理操作、目标特征向量的确定操作以及目标特征向量的更新操作。利用该方法,规避了传统专家经验发中需要大量专家且转嫁能力参差不齐的情况,大幅度提高了企业筛选的智能化程度,通过决策树模型与动态神经网络模型的融合,大幅度提升了企业筛选的整体分类的准确率,保证了准确率处于95%的筛选效果。

[0107] 实施例三

[0108] 图4为本发明实施例三提供的一种企业筛选装置的结构框图,该装置可以采用软件和/或硬件的方式实现,该装置可以配置于设备中,例如典型的是计算机等。如图4所示,该装置具体包括:获取模块31、特征处理模块32、企业分类模块33以及目标企业确定模块34。

[0109] 其中,获取模块31,用于获取至少一个待筛选企业及相应的特征变量集,所述特征变量集包含待筛选企业相对所设定全部特征变量的特征值;

[0110] 特征处理模块32,用于采用预确定的目标决策树结构处理各所述特征变量集,获得各待筛选企业用于企业筛选的目标特征向量;

[0111] 企业分类模块33,用于将各所述目标特征向量输入给定的动态神经网络模型,获得各所述待筛选企业归属的分类集;

[0112] 目标企业确定模块34,用于根据各所述分类集确定满足设定需求的目标企业。

[0113] 本发明实施例三提供的一种企业筛选装置,与现有通过专家经验筛选相比,规避了需要大量专家、时效低以及评估维度片面等问题,实现了招商企业的智能化筛选,且在智

能化筛选中具体通过对待筛选企业所具备全部特征变量的筛选处理确定了有效的特征变量,来作为企业筛选所输入的目标特征向量,从而避免了无效特征变量的干扰,有效保证了企业筛选的准确率。

[0114] 进一步地,该装置还包括了目标结构确定模块,所述目标结构确定模块具体包括:

[0115] 信息初始单元,用于获取包含至少一个样本企业的样本企业集,初始化参与决策的特征变量的决策数量,其中,各所述样本企业对应存在标准特征变量集及标准所属分类;

[0116] 树结构确定单元,用于从所设定全部特征变量中选定包含所述决策数量个特征变量的决策特征集;

[0117] 目标结果确定单元,用于基于所述样本企业集训练形成所述决策特征集对应的决策树结构,并获得各样本企业在所述决策树结构下的目标所属分类。

[0118] 结果准确度确定单元,用于根据各样本企业的标准所属分类及在所述决策树结构下对应的目标所属分类,确定所述决策树结构的分类准确率;

[0119] 循环判定单元,用于当所述分类准确率达到设定阈值时,将所述决策树结构确定为目标决策树结构;否则,将预设定的选定次数加1,返回执行决策特征集的选定操作;直至所述选定次数达到设定次数阈值后,更新所述决策数量并返回决策特征集的选定操作。

[0120] 在上述优化的基础上,所述所设定全部特征变量按照特征价值优先级分布在至少一个特征分类集,相应的,所述树结构确定单元,具体可用于根据所述决策数量及所述特征分类集的个数,确定每个特征分类集中特征变量的选定量;从各所述特征分类集中选取一组由所述选定量个特征变量构成的未选定过的特征组合,基于各所述特征组合形成决策特征集。

[0121] 进一步地,所述决策数量以步长为1更新;相应的,所述所述根据所述决策数量及所述特征分类集的个数,确定每个特征分类集中特征变量的选定量,包括:当所述决策数量整除所述特征分类集的个数时,将整除的商作为每个特征分类集中特征变量的选定量;否则,将所述决策数量与所述特征分类集的商作为每个特征分类集中特征变量当前的选定量;将按优先级由高到低排列的设定个数特征分类集对应的选定量加1,其中,所述设定个数为所述决策数量与所述特征分类集的余数。

[0122] 进一步地,所述目标结果确定单元,具体可用于以待构建树结构第一层的根节点为当前节点,并将所述决策特征集作为当前的候选特征集;针对所述候选特征集中的每个决策特征变量,获取所述样本企业集中各样本企业对应所述决策特征变量的决策特征值;基于各所述决策特征值结合给定的信息增益计算公式,确定所述决策特征变量相对所述当前节点的信息增益率;比对各所述决策特征变量相对所述当前节点的信息增益率,并将最高信息增益率对应的决策特征变量作为所述当前节点的关键特征变量;根据所述关键特征变量从所述决策特征集中确定待构建树结构下一层对应候选特征变量集以及下一层所包含节点的节点数量;将所述节点数量的子节点分别作为新的当前节点,返回执行决策特征值的获取操作直至所述候选特征变量集为空;获得构建的待剪枝决策树结构,并采用各所述样本企业对所述待剪枝决策树结构进行剪枝处理形成精简的决策树结构;根据所述决策树结构中各叶节点对应的数据信息,各样本企业在所述决策树结构下的目标所属分类。

[0123] 在上述实施例的基础上,特征处理模块32具体可用于将各所述特征变量集输入具备所述目标决策树结构的决策树模型;获得从各所述特征变量集筛选的以所述目标决策树

结构排序的目标特征变量序列,各所述目标特征变量序列与所述待筛选企业一一对应;基于各所述目标特征变量序列中对应的特征值,构成各所述待筛选企业用于企业筛选的目标特征向量。

[0124] 进一步地,该装置还优化包括了:文本处理模块和合并处理模块,

[0125] 所述文本处理模块,用于向量化处理各所述待筛选企业对应的企业评价文本,获得各所述企业评价文本的企业评价向量;

[0126] 所述合并处理模块,用于对各所述待筛选企业的目标特征向量及企业评价向量进行合并处理,将合并后形成的向量记为新的目标特征向量。

[0127] 进一步地,该装置还优化包括了:数据处理模块,用于处理所抓取的至少一个企业的原始数据,从处理后的数据中提取所设定全部特征变量的特征值,形成各所述待筛选企业的特征变量集。

[0128] 上述企业筛选装置可执行本发明任意实施例所提供的企业筛选方法,具备执行方法相应的功能模块和有益效果。

[0129] 实施例四

[0130] 图5是本发明实施例四提供的一种企业筛选设备的结构示意图。参考图5,该企业筛选设备包括:处理器810、存储器820、输入装置830以及输出装置840。该企业筛选设备中处理器810的数量可以是一个或者多个,图5中以一个处理器810为例。该企业筛选设备中存储器820的数量可以是一个或者多个,图5中以一个存储器820为例。该企业筛选设备的处理器810、存储器820、输入装置830以及输出装置840可以通过总线或者其他方式连接,图5中通过总线连接为例。实施例中,该企业筛选设备可为具备开发功能的终端设备,比如,台式机、笔记本电脑等。

[0131] 存储器820作为一种计算机可读存储介质,可用于存储软件程序、计算机可执行程序以及模块,如本发明任意实施例所述的企业筛选设备对应的程序指令/模块(例如,企业筛选装置中的获取模块31、特征处理模块32、企业分类模块33以及目标企业确定模块34)。存储器820可主要包括存储程序区和存储数据区,其中,存储程序区可存储操作系统、至少一个功能所需的应用程序;存储数据区可存储根据设备的使用所创建的数据等。此外,存储器820可以包括高速随机存取存储器,还可以包括非易失性存储器,例如至少一个磁盘存储器件、闪存器件、或其他非易失性固态存储器件。在一些实例中,存储器820可进一步包括相对于处理器810远程设置的存储器,这些远程存储器可以通过网络连接至设备。上述网络的实例包括但不限于互联网、企业内部网、局域网、移动通信网及其组合。

[0132] 输入装置830可用于接收输入的数字或者字符信息,以及产生与设备的用户设置以及功能控制有关的键信号输入,还可以是用于获取图像的摄像头以及获取音频数据的拾音设备。输出装置840可以包括扬声器等音频设备。需要说明的是,输入装置830和输出装置840的具体组成可以根据实际情况设定。

[0133] 处理器810通过运行存储在存储器820中的软件程序、指令以及模块,从而执行设备的各种功能应用以及数据处理,即实现上述的企业筛选方法。

[0134] 上述提供的企业筛选设备可用于执行上述任意实施例提供的企业筛选方法,具备相应的功能和有益效果。

[0135] 本发明实施例四还提供一种包含计算机可执行指令的存储介质,所述计算机可执

行指令在由计算机处理器执行时用于执行一种企业筛选方法,包括:

[0136] 获取至少一个待筛选企业及相应的特征变量集,所述特征变量集包含待筛选企业相对所设定全部特征变量的特征值;采用预确定的目标决策树结构处理各所述特征变量集,获得各待筛选企业用于企业筛选的目标特征向量;将各所述目标特征向量输入给定的动态神经网络模型,获得各所述待筛选企业归属的分类集,所述动态神经网络模型采用所述目标决策树结构对应的特征变量进行模型训练;根据各所述分类集确定满足设定需求的目标企业。

[0137] 当然,本发明实施例所提供的一种包含计算机可执行指令的存储介质,其计算机可执行指令不限于如上所述的企业筛选方法操作,还可以执行本发明任意实施例所提供的企业筛选方法中的相关操作,且具备相应的功能和有益效果。

[0138] 通过以上关于实施方式的描述,所属领域的技术人员可以清楚地了解到,本发明可借助软件及必需的通用硬件来实现,当然也可以通过硬件实现,但很多情况下前者是更佳的实施方式。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,该计算机软件产品可以存储在计算机可读存储介质中,如计算机的软盘、只读存储器(Read-Only Memory,ROM)、随机存取存储器(Random Access Memory, RAM)、闪存(FLASH)、硬盘或光盘等,包括若干指令用以使得一台计算机设备(可以是机器人,个人计算机,服务器,或者网络设备等)执行本发明任意实施例所述的企业筛选方法。

[0139] 值得注意的是,上述企业筛选装置中,所包括的各个单元和模块只是按照功能逻辑进行划分的,但并不局限于上述的划分,只要能够实现相应的功能即可;另外,各功能单元的具体名称也只是为了便于相互区分,并不用于限制本发明的保护范围。

[0140] 应当理解,本发明的各部分可以用硬件、软件、固件或它们的组合来实现。在上述实施方式中,多个步骤或方法可以用存储在存储器中且由合适的指令执行系统执行的软件或固件来实现。例如,如果用硬件来实现,和在另一实施方式中一样,可用本领域公知的下列技术中的任一项或他们的组合来实现:具有用于对数据信号实现逻辑功能的逻辑门电路的离散逻辑电路,具有合适的组合逻辑门电路的专用集成电路,可编程门阵列(PGA),现场可编程门阵列(FPGA)等。

[0141] 在本说明书的描述中,参考术语“一个实施例”、“一些实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本发明的至少一个实施例或示例中。在本说明书中,对上述术语的示意性表述不一定指的是相同的实施例或示例。而且,描述的具体特征、结构、材料或者特点可以在任何的一个或多个实施例或示例中以合适的方式结合。

[0142] 以上描述仅为本发明的较佳实施例以及对所运用技术原理的说明。本领域技术人员应当理解,本发明中所涉及的公开范围,并不限于上述技术特征的特定组合而成的技术方案,同时也应涵盖在不脱离上述发明构思的情况下,由上述技术特征或其等同特征进行任意组合而形成的其它技术方案。例如上述特征与本发明中公开的(但不限于)具有类似功能的技术特征进行互相替换而形成的技术方案。

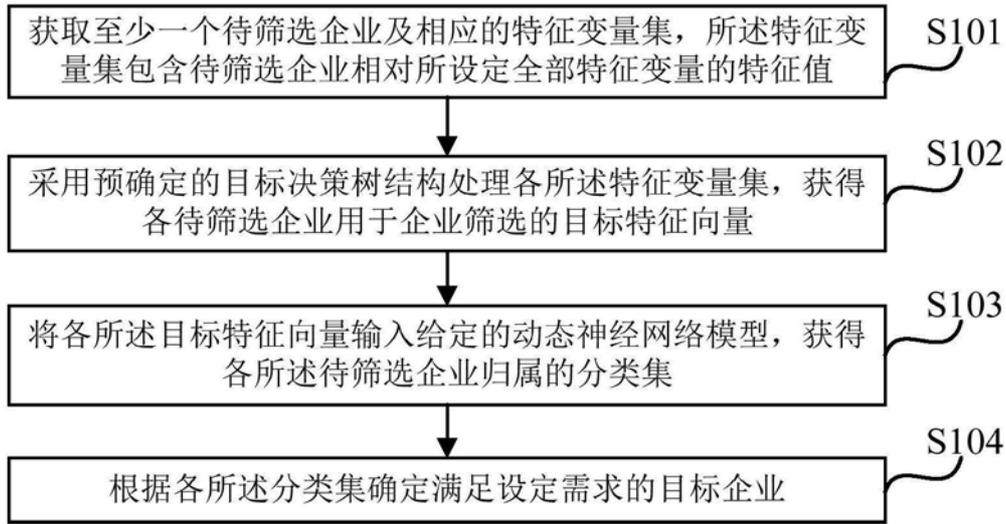


图1

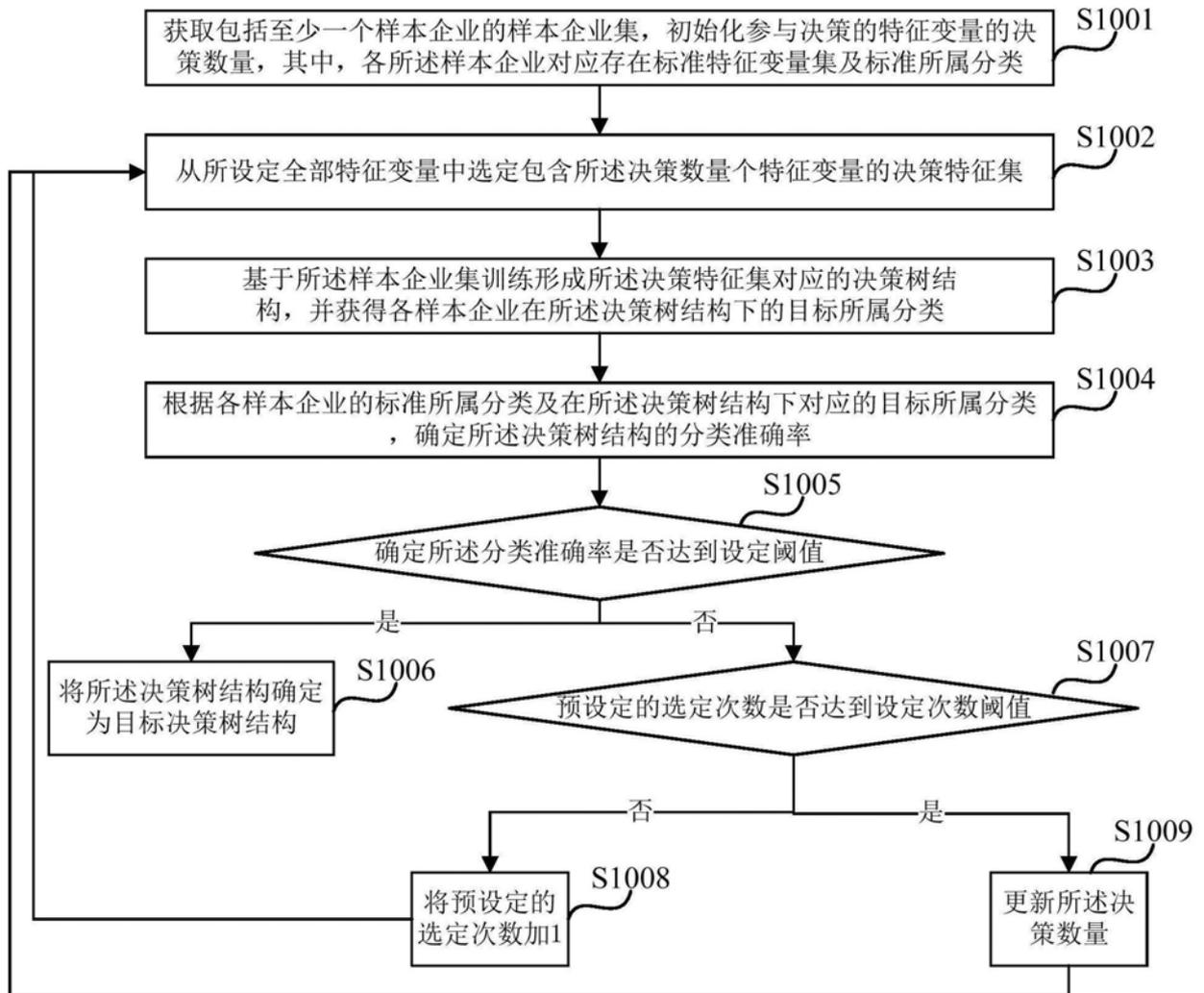


图2

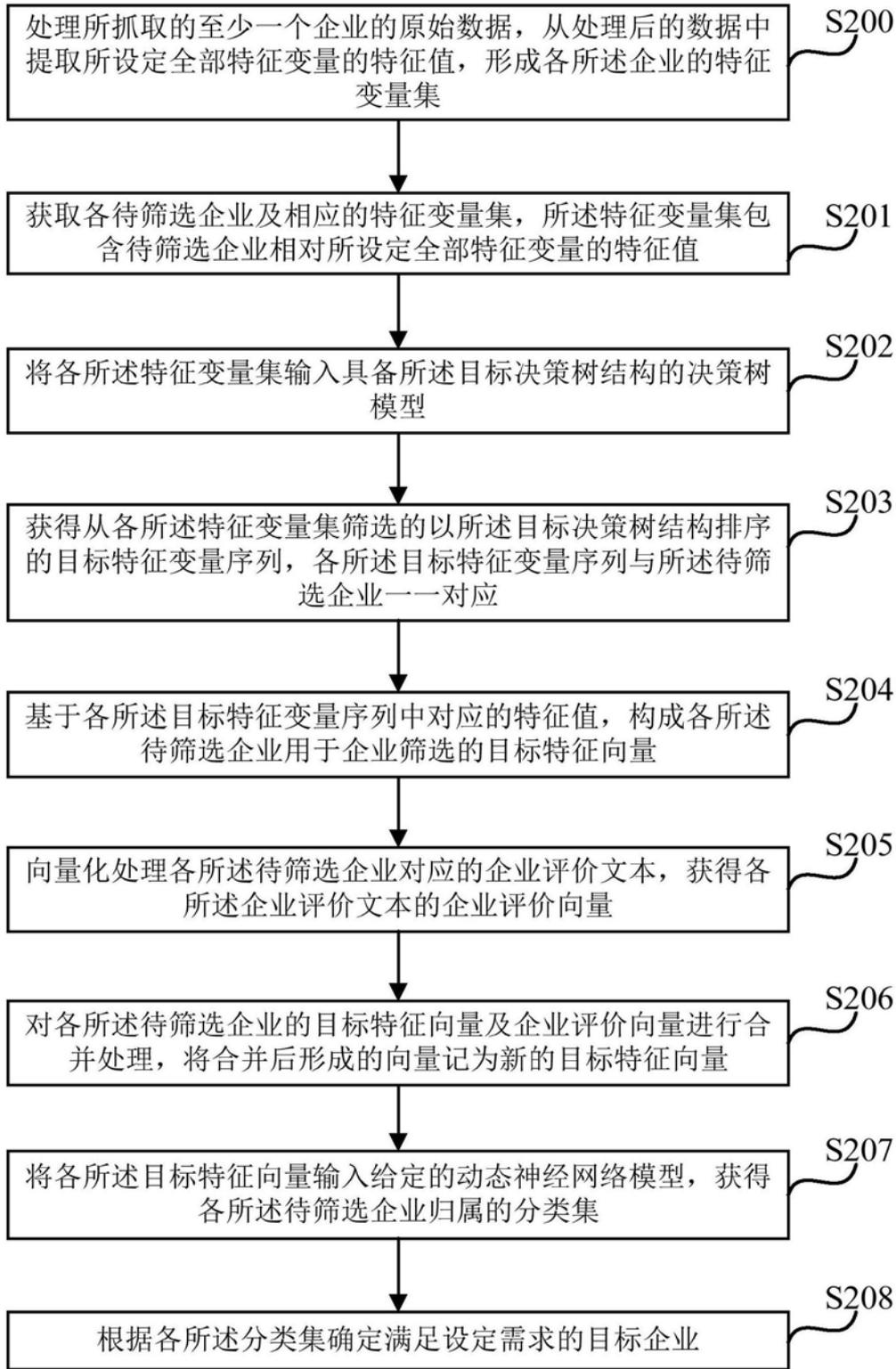


图3

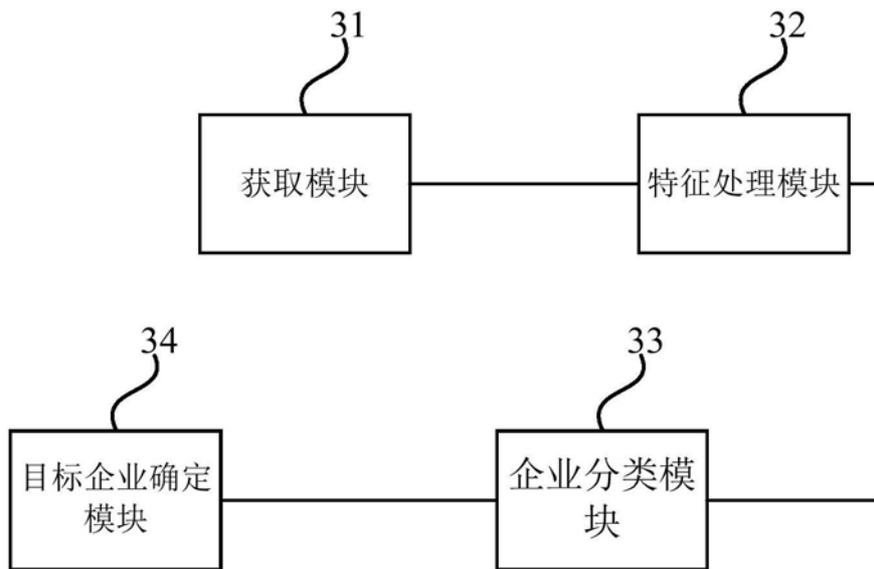


图4

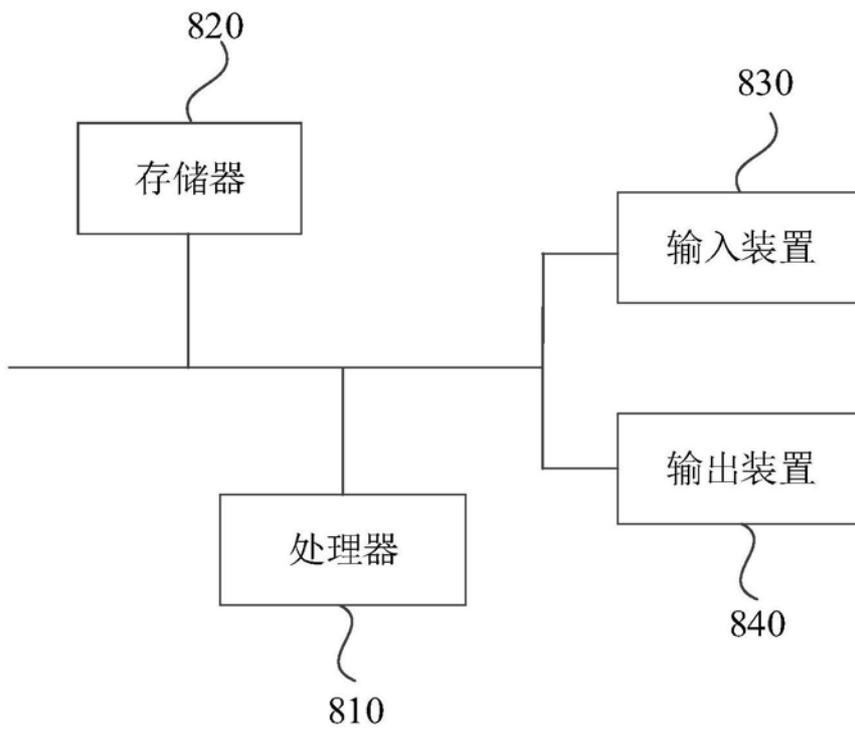


图5