



## (12)发明专利申请

(10)申请公布号 CN 106407999 A

(43)申请公布日 2017.02.15

(21)申请号 201610728180.4

(22)申请日 2016.08.25

(71)申请人 北京物思创想科技有限公司

地址 100085 北京市海淀区上地东路35号  
颐泉汇大厦写字楼303室

(72)发明人 罗远飞 陈雨强

(74)专利代理机构 北京展翼知识产权代理事务  
所(特殊普通合伙) 11452

代理人 屠长存

(51)Int.Cl.

G06K 9/62(2006.01)

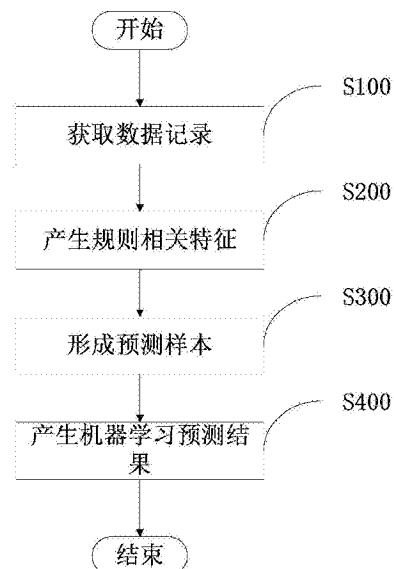
权利要求书2页 说明书25页 附图4页

### (54)发明名称

结合规则来进行机器学习的方法及系统

### (57)摘要

提供了一种结合规则来进行机器学习的方法及系统,所述方法包括:(A)获取数据记录,其中,所述数据记录包括多个属性信息;(B)将关于预测目标的至少一条规则应用于所述多个属性信息,以产生数据记录的规则相关特征;(C)至少基于所述规则相关特征来形成预测样本;以及(D)利用机器学习预测模型,基于预测样本来产生关于预测目标的机器学习预测结果,其中,所述机器学习预测模型被训练为针对预测样本来提供相应的机器学习预测结果。通过所述方法和系统,能够将规则形成为参与机器学习的规则相关特征,从而有效地将规则融入到机器学习的过程中,由此改进预测效果。



1. 一种结合规则来进行机器学习的方法,包括:

(A) 获取数据记录,其中,所述数据记录包括多个属性信息;

(B) 将关于预测目标的至少一条规则应用于所述多个属性信息,以产生数据记录的规则相关特征;

(C) 至少基于所述规则相关特征来形成预测样本;以及

(D) 利用机器学习预测模型,基于预测样本来产生关于预测目标的机器学习预测结果,其中,所述机器学习预测模型被训练为针对预测样本来提供相应的机器学习预测结果。

2. 如权利要求1所述的方法,其中,规则相关特征包括规则预测特征和/或规则描述特征,其中,在步骤(B)中,基于所述数据记录依据所述至少一条规则而获得的规则预测结果来产生规则预测特征,并且/或者,基于所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立来产生规则描述特征。

3. 如权利要求2所述的方法,其中,在步骤(B)中,通过将指示所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值分别乘以与所述规则描述特征相应的规则的权重来产生规则描述特征;

或者,在步骤(B)中,将指示所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值作为规则描述特征,并且,所述至少一条规则之中的各条规则的权重分别作为机器学习预测模型的规则描述特征系数的初始值,

其中,所述权重表示相应规则的确定性。

4. 如权利要求2所述的方法,其中,在步骤(B)中,通过将指示所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值分别乘以与所述规则描述特征相应的规则的结论值来产生规则描述特征;

或者,在步骤(B)中,将指示所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值作为规则描述特征,并且,所述至少一条规则之中的各条规则的结论值分别作为机器学习预测模型的规则描述特征系数的初始值,

其中,所述结论值表示相应规则的条件成立时所导致的结果。

5. 如权利要求2所述的方法,其中,在步骤(B)中,通过将指示所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值分别乘以与所述规则描述特征相应的规则的权重与结论值的乘积来产生规则描述特征;

或者,在步骤(B)中,将指示所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值作为规则描述特征,并且,所述至少一条规则之中的各条规则的权重与结论值的乘积分别作为机器学习预测模型的规则描述特征系数的初始值,

其中,所述权重表示相应规则的确定性,所述结论值表示相应规则的条件成立时所导致的结果。

6. 一种结合规则来进行机器学习的方法,包括:

(A) 获取历史数据记录,其中,所述历史数据记录包括多个属性信息和作为预测目标实际值的标记;

(B) 将关于预测目标的至少一条规则应用于所述多个属性信息,以产生历史数据记录的规则相关特征;

(C) 至少基于所述规则相关特征和标记来形成训练样本;以及

(D) 基于训练样本来训练机器学习预测模型,其中,所述机器学习模型用于针对新的数据记录来提供关于预测目标的机器学习预测结果。

7. 一种结合规则来进行机器学习的系统,包括:

数据记录获取装置,用于获取数据记录,其中,所述数据记录包括多个属性信息;

规则相关特征产生装置,用于将关于预测目标的至少一条规则应用于所述多个属性信息,以产生数据记录的规则相关特征;

预测样本产生装置,用于至少基于所述规则相关特征来形成预测样本;以及

机器学习预测装置,用于利用机器学习预测模型,基于预测样本来产生关于预测目标的机器学习预测结果,其中,所述机器学习预测模型被训练为针对预测样本来提供相应的机器学习预测结果。

8. 一种结合规则来进行机器学习的系统,包括:

历史数据记录获取装置,用于获取历史数据记录,其中,所述历史数据记录包括多个属性信息和作为预测目标实际值的标记;

规则相关特征产生装置,用于将关于预测目标的至少一条规则应用于所述多个属性信息,以产生历史数据记录的规则相关特征;

训练样本产生装置,用于至少基于所述规则相关特征和标记来形成训练样本;以及

机器学习模型训练装置,用于基于训练样本来训练机器学习预测模型,其中,所述机器学习模型用于针对新的数据记录来提供关于预测目标的机器学习预测结果。

9. 一种结合规则来进行机器学习的计算装置,包括存储部件和处理器,存储部件中存储有计算机可执行指令集合,当所述计算机可执行指令集合被所述处理器执行时,执行下述步骤:

(A) 获取数据记录,其中,所述数据记录包括多个属性信息;

(B) 将关于预测目标的至少一条规则应用于所述多个属性信息,以产生数据记录的规则相关特征;

(C) 至少基于所述规则相关特征来形成预测样本;以及

(D) 利用机器学习预测模型,基于预测样本来产生关于预测目标的机器学习预测结果,其中,所述机器学习预测模型被训练为针对预测样本来提供相应的机器学习预测结果。

10. 一种结合规则来进行机器学习的计算装置,包括存储部件和处理器,存储部件中存储有计算机可执行指令集合,当所述计算机可执行指令集合被所述处理器执行时,执行下述步骤:

(A) 获取历史数据记录,其中,所述历史数据记录包括多个属性信息和作为预测目标实际值的标记;

(B) 将关于预测目标的至少一条规则应用于所述多个属性信息,以产生历史数据记录的规则相关特征;

(C) 至少基于所述规则相关特征和标记来形成训练样本;以及

(D) 基于训练样本来训练机器学习预测模型,其中,所述机器学习模型用于针对新的数据记录来提供关于预测目标的机器学习预测结果。

## 结合规则来进行机器学习的方法及系统

### 技术领域

[0001] 本发明的示例性实施例总体说来涉及人工智能领域,更具体地说,涉及一种结合规则来进行机器学习(例如,训练和预测)的方法及系统。

### 背景技术

[0002] 近来,机器学习技术与传统的规则(Rule)系统相比,在人工智能领域得到了更为普遍的应用。这是因为,随着应用场景的复杂化,需要归纳和开发的规则数量变得越来越大,有些时候,数据的增长速度导致人们已经无法掌握数据变化所带来的规则改变。

[0003] 相应地,人们更倾向于使用机器学习技术来解决问题。然而,很多机器学习算法均为黑盒,产生的机器学习模型往往很难清楚地表达某个具体策略的原因;另一方面,即使在实践中已经发现某个因素在特定场景下具有关键作用,也很难将这样的发现有效地直接施加到机器学习系统中。

[0004] 具体说来,在公开号为US20160171386号的美国专利申请中,介绍了一种用于观点挖掘的系统和方法,其中,基于规则的系统可作为情绪检测模块,基于机器学习的系统可与所述情绪检测模块进行通信并处理由情绪检测模块提供的数据。

[0005] 此外,在公开号为CN105721194A的中国专利申请中,公开了一种利用大数据实时处理、机器学习技术等技术,实现全自动的运营商网络故障定位功能的方案。将运维人员的经验固化到系统中,使系统具备基础的判断智能。

[0006] 另外,在公开号为CN105320960A的中国专利申请中,公开了一种基于投票的跨语言主客观情感分类方法,包括如下步骤:S1、根据源语言的情感词典构建目标语言的情感词典;S2、采用规则算法、机器翻译与统计机器学习相结合的算法、极性特征值计算算法三种算法分别从待标注文本中的句子中提取词,根据构建的目标语言的情感词典,判断词的情感极性,进而进行句子的主客观性质判断;S3、获得根据三种算法得到的句子的主客观性质的判断结果,通过投票对句子的主客观性质进行判定。

[0007] 可以看出,在已有的解决方案中,规则系统和机器学习系统可顺序连接,即,通过规则系统对机器学习系统的数据样本进行预处理,或对机器学习系统的预测结果进行修正;或者,规则系统和机器学习系统可分别进行预测,并在两者的判断结果中择一使用。上述方式中,规则系统与机器学习系统均独立工作,使得规则难以有效地融入到机器学习中。

### 发明内容

[0008] 本发明的示例性实施例旨在克服规则难以有效地融入机器学习的缺陷。

[0009] 根据本发明的示例性实施例,提供一种结合规则来进行机器学习的方法,包括:(A)获取数据记录,其中,所述数据记录包括多个属性信息;(B)将关于预测目标的至少一条规则应用于所述多个属性信息,以产生数据记录的规则相关特征;(C)至少基于所述规则相关特征来形成预测样本;以及(D)利用机器学习预测模型,基于预测样本来产生关于预测目标的机器学习预测结果,其中,所述机器学习预测模型被训练为针对预测样本来提供相应

的机器学习预测结果。

[0010] 可选地,在所述方法中,规则相关特征包括规则预测特征和/或规则描述特征,其中,在步骤(B)中,基于所述数据记录依据所述至少一条规则而获得的规则预测结果来产生规则预测特征,并且/或者,基于所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立来产生规则描述特征。

[0011] 可选地,在所述方法中,在步骤(B)中,通过将指示所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值分别乘以与所述规则描述特征相应的规则的权重来产生规则描述特征;或者,在步骤(B)中,将指示所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值作为规则描述特征,并且,所述至少一条规则之中的各条规则的权重分别作为机器学习预测模型的规则描述特征系数的初始值,其中,所述权重表示相应规则的确定性。

[0012] 可选地,在所述方法中,所述权重通过人为指定和/或通过规则学习引擎来设置,其中,所述规则学习引擎用于基于规则训练样本来学习所述至少一条规则之中的各条规则的权重。

[0013] 可选地,在所述方法中,在步骤(B)中,通过将指示所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值分别乘以与所述规则描述特征相应的规则的结论值来产生规则描述特征;或者,在步骤(B)中,将指示所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值作为规则描述特征,并且,所述至少一条规则之中的各条规则的结论值分别作为机器学习预测模型的规则描述特征系数的初始值,其中,所述结论值表示相应规则的条件成立时所导致的结果。

[0014] 可选地,在所述方法中,在所述结果与预测目标正相关时,所述结论值为正值,并且,在所述结果与预测目标负相关时,所述结论值为负值。

[0015] 可选地,在所述方法中,在步骤(B)中,通过将指示所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值分别乘以与所述规则描述特征相应的规则的权重与结论值的乘积来产生规则描述特征;或者,在步骤(B)中,将指示所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值作为规则描述特征,并且,所述至少一条规则之中的各条规则的权重与结论值的乘积分别作为机器学习预测模型的规则描述特征系数的初始值,其中,所述权重表示相应规则的确定性,所述结论值表示相应规则的条件成立时所导致的结果。

[0016] 可选地,在所述方法中,机器学习预测模型的规则描述特征系数的初始值被乘以根据机器学习预测模型的算法而设置的系数修正值。

[0017] 可选地,所述方法还包括:(E)将机器学习预测结果与所述数据记录依据所述至少一条规则而获得的规则预测结果进行融合,以得到与预测样本相应的融合预测结果。

[0018] 可选地,所述方法还包括:(F)将数据记录与机器学习预测结果或融合预测结果组合为规则训练样本。

[0019] 可选地,在所述方法中,所述至少一条规则之中的各条规则的权重基于机器学习预测模型的规则描述特征系数的更新值来设置。

[0020] 可选地,在所述方法中,规则学习引擎基于马尔可夫逻辑网络。

[0021] 可选地,在步骤(B)之前,所述方法还包括:(G)获取所述关于预测目标的至少一条

规则。

[0022] 根据本发明的另一示例性实施例，提供一种结合规则来进行机器学习的方法，包括：(A) 获取历史数据记录，其中，所述历史数据记录包括多个属性信息和作为预测目标实际值的标记；(B) 将关于预测目标的至少一条规则应用于所述多个属性信息，以产生历史数据记录的规则相关特征；(C) 至少基于所述规则相关特征和标记来形成训练样本；以及(D) 基于训练样本来训练机器学习预测模型，其中，所述机器学习模型用于针对新的数据记录来提供关于预测目标的机器学习预测结果。

[0023] 可选地，在所述方法中，规则相关特征包括规则预测特征和/或规则描述特征，其中，在步骤(B)中，基于所述历史数据记录依据所述至少一条规则而获得的规则预测结果来产生规则预测特征，并且/或者，基于所述历史数据记录针对所述至少一条规则之中的各条规则的条件是否成立来产生规则描述特征。

[0024] 可选地，在所述方法中，在步骤(B)中，通过将指示所述历史数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值分别乘以与所述规则描述特征相应的规则的权重来产生规则描述特征；或者，在步骤(B)中，将指示所述历史数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值作为规则描述特征，并且，在步骤(D)中，将所述至少一条规则之中的各条规则的权重分别作为机器学习预测模型的规则描述特征系数的初始值，其中，所述权重表示相应规则的确定性。

[0025] 可选地，在所述方法中，所述权重通过人为指定和/或通过规则学习引擎来设置，其中，所述规则学习引擎用于基于规则训练样本来学习所述至少一条规则之中的各条规则的权重。

[0026] 可选地，在所述方法中，在步骤(B)中，通过将指示所述历史数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值分别乘以与所述规则描述特征相应的规则的结论值来产生规则描述特征；或者，在步骤(B)中，将指示所述历史数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值作为规则描述特征，并且，在步骤(D)中，将所述至少一条规则之中的各条规则的结论值分别作为机器学习预测模型的规则描述特征系数的初始值，其中，所述结论值表示相应规则的条件成立时所导致的结果。

[0027] 可选地，在所述方法中，在所述结果与预测目标正相关时，所述结论值为正值，并且，在所述结果与预测目标负相关时，所述结论值为负值。

[0028] 可选地，在所述方法中，在步骤(B)中，通过将指示所述历史数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值分别乘以与所述规则描述特征相应的规则的权重与结论值的乘积来产生规则描述特征；或者，在步骤(B)中，将指示所述历史数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值作为规则描述特征，并且，在步骤(D)中，将所述至少一条规则之中的各条规则的权重与结论值的乘积分别作为机器学习预测模型的规则描述特征系数的初始值，其中，所述权重表示相应规则的确定性，所述结论值表示相应规则的条件成立时所导致的结果。

[0029] 可选地，在所述方法中，在步骤(D)中，将机器学习预测模型的规则描述特征系数的初始值乘以根据机器学习预测模型的算法而设置的系数修正值。

[0030] 可选地，在所述方法中，所述至少一条规则之中的各条规则的权重基于机器学习预测模型的规则描述特征系数的更新值来设置。

- [0031] 可选地,在所述方法中,规则学习引擎基于马尔可夫逻辑网络。
- [0032] 可选地,在步骤(B)之前,所述方法还包括:(G)获取所述关于预测目标的至少一条规则。
- [0033] 根据本发明的另一示例性实施例,提供一种结合规则来进行机器学习的系统,包括:数据记录获取装置,用于获取数据记录,其中,所述数据记录包括多个属性信息;规则相关特征产生装置,用于将关于预测目标的至少一条规则应用于所述多个属性信息,以产生数据记录的规则相关特征;预测样本产生装置,用于至少基于所述规则相关特征来形成预测样本;以及机器学习预测装置,用于利用机器学习预测模型,基于预测样本来产生关于预测目标的机器学习预测结果,其中,所述机器学习预测模型被训练为针对预测样本来提供相应的机器学习预测结果。
- [0034] 可选地,在所述系统中,规则相关特征包括规则预测特征和/或规则描述特征,其中,规则相关特征产生装置基于所述数据记录依据所述至少一条规则而获得的规则预测结果来产生规则预测特征,并且/或者,基于所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立来产生规则描述特征。
- [0035] 可选地,在所述系统中,规则相关特征产生装置通过将指示所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值分别乘以与所述规则描述特征相应的规则的权重来产生规则描述特征;或者,规则相关特征产生装置将指示所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值作为规则描述特征,并且,所述至少一条规则之中的各条规则的权重分别作为机器学习预测模型的规则描述特征系数的初始值,其中,所述权重表示相应规则的确定性。
- [0036] 可选地,在所述系统中,所述权重通过人为指定和/或通过规则学习引擎来设置,其中,所述规则学习引擎用于基于规则训练样本来学习所述至少一条规则之中的各条规则的权重。
- [0037] 可选地,在所述系统中,规则相关特征产生装置通过将指示所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值分别乘以与所述规则描述特征相应的规则的结论值来产生规则描述特征;或者,规则相关特征产生装置将指示所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值作为规则描述特征,并且,所述至少一条规则之中的各条规则的结论值分别作为机器学习预测模型的规则描述特征系数的初始值,其中,所述结论值表示相应规则的条件成立时所导致的结果。
- [0038] 可选地,在所述系统中,在所述结果与预测目标正相关时,所述结论值为正值,并且,在所述结果与预测目标负相关时,所述结论值为负值。
- [0039] 可选地,在所述系统中,规则相关特征产生装置通过将指示所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值分别乘以与所述规则描述特征相应的规则的权重与结论值的乘积来产生规则描述特征;或者,规则相关特征产生装置将指示所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值作为规则描述特征,并且,所述至少一条规则之中的各条规则的权重与结论值的乘积分别作为机器学习预测模型的规则描述特征系数的初始值,其中,所述权重表示相应规则的确定性,所述结论值表示相应规则的条件成立时所导致的结果。
- [0040] 可选地,在所述系统中,机器学习预测模型的规则描述特征系数的初始值被乘以

根据机器学习预测模型的算法而设置的系数修正值。

[0041] 可选地，所述系统还包括：融合装置，用于将机器学习预测结果与所述数据记录依据所述至少一条规则而获得的规则预测结果进行融合，以得到与预测样本相应的融合预测结果。

[0042] 可选地，在所述系统中，机器学习预测装置还将数据记录与机器学习预测结果组合为规则训练样本；或者，融合装置还将数据记录与融合预测结果组合为规则训练样本。

[0043] 可选地，在所述系统中，所述至少一条规则之中的各条规则的权重基于机器学习预测模型的规则描述特征系数的更新值来设置。

[0044] 可选地，在所述系统中，规则学习引擎基于马尔可夫逻辑网络。

[0045] 可选地，所述系统还包括：规则获取装置，用于获取所述关于预测目标的至少一条规则。

[0046] 根据本发明的另一示例性实施例，提供一种结合规则来进行机器学习的系统，包括：历史数据记录获取装置，用于获取历史数据记录，其中，所述历史数据记录包括多个属性信息和作为预测目标实际值的标记；规则相关特征产生装置，用于将关于预测目标的至少一条规则应用于所述多个属性信息，以产生历史数据记录的规则相关特征；训练样本产生装置，用于至少基于所述规则相关特征和标记来形成训练样本；以及机器学习模型训练装置，用于基于训练样本来训练机器学习预测模型，其中，所述机器学习模型用于针对新的数据记录来提供关于预测目标的机器学习预测结果。

[0047] 可选地，在所述系统中，规则相关特征包括规则预测特征和/或规则描述特征，其中，规则相关特征产生装置基于所述历史数据记录依据所述至少一条规则而获得的规则预测结果来产生规则预测特征，并且/或者，基于所述历史数据记录针对所述至少一条规则之中的各条规则的条件是否成立来产生规则描述特征。

[0048] 可选地，在所述系统中，规则相关特征产生装置通过将指示所述历史数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值分别乘以与所述规则描述特征相应的规则的权重来产生规则描述特征；或者，规则相关特征产生装置将指示所述历史数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值作为规则描述特征，并且，机器学习模型训练装置将所述至少一条规则之中的各条规则的权重分别作为机器学习预测模型的规则描述特征系数的初始值，其中，所述权重表示相应规则的确定性。

[0049] 可选地，在所述系统中，所述权重通过人为指定和/或通过规则学习引擎来设置，其中，所述规则学习引擎用于基于规则训练样本来学习所述至少一条规则之中的各条规则的权重。

[0050] 可选地，在所述系统中，规则相关特征产生装置通过将指示所述历史数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值分别乘以与所述规则描述特征相应的规则的结论值来产生规则描述特征；或者，规则相关特征产生装置将指示所述历史数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值作为规则描述特征，并且，机器学习模型训练装置将所述至少一条规则之中的各条规则的结论值分别作为机器学习预测模型的规则描述特征系数的初始值，其中，所述结论值表示相应规则的条件成立时所导致的结果。

[0051] 可选地,在所述系统中,在所述结果与预测目标正相关时,所述结论值为正值,并且,在所述结果与预测目标负相关时,所述结论值为负值。

[0052] 可选地,在所述系统中,规则相关特征产生装置通过将指示所述历史数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值分别乘以与所述规则描述特征相应的规则的权重与结论值的乘积来产生规则描述特征;或者,规则相关特征产生装置将指示所述历史数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值作为规则描述特征,并且,机器学习模型训练装置将所述至少一条规则之中的各条规则的权重与结论值的乘积分别作为机器学习预测模型的规则描述特征系数的初始值,其中,所述权重表示相应规则的确定性,所述结论值表示相应规则的条件成立时所导致的结果。

[0053] 可选地,在所述系统中,机器学习模型训练装置将机器学习预测模型的规则描述特征系数的初始值乘以根据机器学习预测模型的算法而设置的系数修正值。

[0054] 可选地,在所述系统中,所述至少一条规则之中的各条规则的权重基于机器学习预测模型的规则描述特征系数的更新值来设置。

[0055] 可选地,在所述系统中,规则学习引擎基于马尔可夫逻辑网络。

[0056] 可选地,所述系统还包括:规则获取装置,用于获取所述关于预测目标的至少一条规则。

[0057] 根据本发明的另一示例性实施例,提供一种结合规则来进行机器学习的计算装置,包括存储部件和处理器,存储部件中存储有计算机可执行指令集合,当所述计算机可执行指令集合被所述处理器执行时,执行下述步骤:(A)获取数据记录,其中,所述数据记录包括多个属性信息;(B)将关于预测目标的至少一条规则应用于所述多个属性信息,以产生数据记录的规则相关特征;(C)至少基于所述规则相关特征来形成预测样本;以及(D)利用机器学习预测模型,基于预测样本来产生关于预测目标的机器学习预测结果,其中,所述机器学习预测模型被训练为针对预测样本来提供相应的机器学习预测结果。

[0058] 可选地,在所述计算装置中,规则相关特征包括规则预测特征和/或规则描述特征,其中,在步骤(B)中,基于所述数据记录依据所述至少一条规则而获得的规则预测结果来产生规则预测特征,并且/或者,基于所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立来产生规则描述特征。

[0059] 可选地,在所述计算装置中,在步骤(B)中,通过将指示所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值分别乘以与所述规则描述特征相应的规则的权重来产生规则描述特征;或者,在步骤(B)中,将指示所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值作为规则描述特征,并且,所述至少一条规则之中的各条规则的权重分别作为机器学习预测模型的规则描述特征系数的初始值,其中,所述权重表示相应规则的确定性。

[0060] 可选地,在所述计算装置中,所述权重通过人为指定和/或通过规则学习引擎来设置,其中,所述规则学习引擎用于基于规则训练样本来学习所述至少一条规则之中的各条规则的权重。

[0061] 可选地,在所述计算装置中,在步骤(B)中,通过将指示所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值分别乘以与所述规则描述特征相

的规则的结论值来产生规则描述特征；或者，在步骤(B)中，将指示所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值作为规则描述特征，并且，所述至少一条规则之中的各条规则的结论值分别作为机器学习预测模型的规则描述特征系数的初始值，其中，所述结论值表示相应规则的条件成立时所导致的结果。

[0062] 可选地，在所述计算装置中，在所述结果与预测目标正相关时，所述结论值为正值，并且，在所述结果与预测目标负相关时，所述结论值为负值。

[0063] 可选地，在所述计算装置中，在步骤(B)中，通过将指示所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值分别乘以与所述规则描述特征相应的规则的权重与结论值的乘积来产生规则描述特征；或者，在步骤(B)中，将指示所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值作为规则描述特征，并且，所述至少一条规则之中的各条规则的权重与结论值的乘积分别作为机器学习预测模型的规则描述特征系数的初始值，其中，所述权重表示相应规则的确定性，所述结论值表示相应规则的条件成立时所导致的结果。

[0064] 可选地，在所述计算装置中，机器学习预测模型的规则描述特征系数的初始值被乘以根据机器学习预测模型的算法而设置的系数修正值。

[0065] 可选地，在所述计算装置中，当所述计算机可执行指令集合被所述处理器执行时，还执行下述步骤：(E) 将机器学习预测结果与所述数据记录依据所述至少一条规则而获得的规则预测结果进行融合，以得到与预测样本相应的融合预测结果。

[0066] 可选地，在所述计算装置中，当所述计算机可执行指令集合被所述处理器执行时，还执行下述步骤：(F) 将数据记录与机器学习预测结果或融合预测结果组合为规则训练样本。

[0067] 可选地，在所述计算装置中，所述至少一条规则之中的各条规则的权重基于机器学习预测模型的规则描述特征系数的更新值来设置。

[0068] 可选地，在所述计算装置中，规则学习引擎基于马尔可夫逻辑网络。

[0069] 可选地，在所述计算装置中，当所述计算机可执行指令集合被所述处理器执行时，在步骤(B)之前，还执行下述步骤：(G) 获取所述关于预测目标的至少一条规则。

[0070] 根据本发明的另一示例性实施例，提供一种结合规则来进行机器学习的计算装置，包括存储部件和处理器，存储部件中存储有计算机可执行指令集合，当所述计算机可执行指令集合被所述处理器执行时，执行下述步骤：(A) 获取历史数据记录，其中，所述历史数据记录包括多个属性信息和作为预测目标实际值的标记；(B) 将关于预测目标的至少一条规则应用于所述多个属性信息，以产生历史数据记录的规则相关特征；(C) 至少基于所述规则相关特征和标记来形成训练样本；以及(D) 基于训练样本来训练机器学习预测模型，其中，所述机器学习模型用于针对新的数据记录来提供关于预测目标的机器学习预测结果。

[0071] 可选地，在所述计算装置中，规则相关特征包括规则预测特征和/或规则描述特征，其中，在步骤(B)中，基于所述历史数据记录依据所述至少一条规则而获得的规则预测结果来产生规则预测特征，并且/或者，基于所述历史数据记录针对所述至少一条规则之中的各条规则的条件是否成立来产生规则描述特征。

[0072] 可选地，在所述计算装置中，在步骤(B)中，通过将指示所述历史数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值分别乘以与所述规则描述特征

相应的规则的权重来产生规则描述特征；或者，在步骤(B)中，将指示所述历史数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值作为规则描述特征，并且，在步骤(D)中，将所述至少一条规则之中的各条规则的权重分别作为机器学习预测模型的规则描述特征系数的初始值，其中，所述权重表示相应规则的确定性。

[0073] 可选地，在所述计算装置中，所述权重通过人为指定和/或通过规则学习引擎来设置，其中，所述规则学习引擎用于基于规则训练样本来学习所述至少一条规则之中的各条规则的权重。

[0074] 可选地，在所述计算装置中，在步骤(B)中，通过将指示所述历史数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值分别乘以与所述规则描述特征相应的规则的结论值来产生规则描述特征；或者，在步骤(B)中，将指示所述历史数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值作为规则描述特征，并且，在步骤(D)中，将所述至少一条规则之中的各条规则的结论值分别作为机器学习预测模型的规则描述特征系数的初始值，其中，所述结论值表示相应规则的条件成立时所导致的结果。

[0075] 可选地，在所述计算装置中，在所述结果与预测目标正相关时，所述结论值为正值，并且，在所述结果与预测目标负相关时，所述结论值为负值。

[0076] 可选地，在所述计算装置中，在步骤(B)中，通过将指示所述历史数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值分别乘以与所述规则描述特征相应的规则的权重与结论值的乘积来产生规则描述特征；或者，在步骤(B)中，将指示所述历史数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值作为规则描述特征，并且，在步骤(D)中，将所述至少一条规则之中的各条规则的权重与结论值的乘积分别作为机器学习预测模型的规则描述特征系数的初始值，其中，所述权重表示相应规则的确定性，所述结论值表示相应规则的条件成立时所导致的结果。

[0077] 可选地，在所述计算装置中，在步骤(D)中，将机器学习预测模型的规则描述特征系数的初始值乘以根据机器学习预测模型的算法而设置的系数修正值。

[0078] 可选地，在所述计算装置中，所述至少一条规则之中的各条规则的权重基于机器学习预测模型的规则描述特征系数的更新值来设置。

[0079] 可选地，在所述计算装置中，规则学习引擎基于马尔可夫逻辑网络。

[0080] 可选地，在所述计算装置中，当所述计算机可执行指令集合被所述处理器执行时，还执行下述步骤：(E)获取所述关于预测目标的至少一条规则。

[0081] 在根据本发明示例性实施例的结合规则来进行机器学习的方法及系统中，能够将规则形成为参与机器学习的规则相关特征，从而有效地将规则融入到机器学习的过程中，由此改进机器学习的效果。

## 附图说明

[0082] 从下面结合附图对本发明实施例的详细描述中，本发明的这些和/或其他方面和优点将变得更加清楚并更容易理解，其中：

[0083] 图1示出根据本发明示例性实施例的结合规则来进行机器学习的系统的框图；

[0084] 图2示出根据本发明另一示例性实施例的结合规则来进行机器学习的系统的框

图；

- [0085] 图3示出根据本发明示例性实施例的结合规则来进行机器学习的方法的流程图；
- [0086] 图4示出根据本发明示例性实施例的马尔可夫逻辑网络的示例；
- [0087] 图5示出根据本发明另一示例性实施例的结合规则来进行机器学习的方法的流程图；
- [0088] 图6示出本发明另一示例性实施例的结合规则来进行机器学习的系统的框图；以及
- [0089] 图7示出根据本发明另一示例性实施例的结合规则来进行机器学习的方法的流程图。

### 具体实施方式

[0090] 为了使本领域技术人员更好地理解本发明，下面结合附图和具体实施方式对本发明的示例性实施例作进一步详细说明。

[0091] 在本发明的示例性实施例中，通过以下方式进行机器学习：通过将关于预测目标的规则应用于预测或训练用的数据记录的各个属性信息，从而产生将参与机器学习的规则相关特征；利用至少涵盖规则相关特征的机器学习样本（例如，预测样本或训练样本）来执行机器学习模型的预测或训练。

[0092] 这里，机器学习是人工智能研究发展到一定阶段的必然产物，其致力于通过计算的手段，利用经验来改善系统自身的性能。在计算机系统中，“经验”通常以“数据”形式存在，通过机器学习算法，可从数据中产生“模型”，也就是说，将经验数据提供给机器学习算法，就能基于这些经验数据产生模型，在面对新的情况时，模型会提供相应的判断，即，预测结果。应注意，本发明对具体的机器学习算法并不进行特定限制。

[0093] 这里，规则通常是指语义明确、能描述数据分布所隐含的客观规律或领域概念、可写成“若……，则……”形式的逻辑规则。形式化地看，一条规则形如：

[0094]  $\oplus \leftarrow f_1 \wedge f_2 \wedge \dots \wedge f_L$

[0095] 其中，逻辑蕴含符号“ $\leftarrow$ ”右边部分称为“逻辑体（body）”，表示该条规则的前提（即，规则的条件），左边部分称为“规则头（head）”，表示该条规则的条件成立时所导致的结果（即，规则的结论值）。规则体是由逻辑文字（literal） $f_k$ 组成的合取式（conjunction），其中，合取符号“ $\wedge$ ”用来表示“并且”，每个文字 $f_k$ 都是对示例属性进行检验的布尔表达式，例如“（色泽=乌黑）”或“ $\neg$ （根蒂=硬挺）”。L是规则体中逻辑文字的个数，成为规则的长度，规则头的“ $\oplus$ ”可以是逻辑文字，一般用来表示规则所判定的目标类别或概念，例如“好瓜”，此外，规则头也可以表示量化的判定结果，例如“瓜的具体甜度”，上述逻辑规则也被称为“if-then规则”。

[0096] 以关于西瓜好坏的规则为例：

[0097] 规则1：好瓜 $\leftarrow$ （根蒂=蜷缩） $\wedge$ （脐部=凹陷）；

[0098] 规则2： $\neg$ 好瓜 $\leftarrow$ （纹理=模糊）。

[0099] 规则1的长度为2，它通过判断两个逻辑文字的赋值（valuation）来对示例进行判别，符合该规则的样本称为被该规则“覆盖（cover）”。需注意的是，被规则1覆盖的样本是好瓜，但没被规则1覆盖的未必不是好瓜；只有被规则2这样以“ $\neg$ 好瓜”为头的规则覆盖的才

不是好瓜。

[0100] 根据本发明的示例性实施例，规则的规则头可与预测目标直接或间接相关(例如，正相关或负相关)，相应地，规则体则是针对数据记录的各个属性信息的具体检验。例如，在预测目标涉及信用卡交易的欺诈审核时，规则头可指示对于信用卡交易是否涉嫌欺诈的判断结果，规则体可包括对关于该信用卡交易情况的属性信息或其他相关信息的具体检验。在存在关于预测目标的多条规则的情况下，可不必要求每条规则的规则头均直接与预测目标相关，而可以是所述多条规则互相关联从而整体上与预测目标相关。

[0101] 相应地，在本发明的示例性实施例中，可将上述关于预测目标的规则有效地融入机器学习模型的预测或训练样本中，从而取得较好的机器学习效果。

[0102] 以下将结合附图来详细描述根据本发明示例性实施例的结合规则来进行机器学习的具体方案。

[0103] 图1示出根据本发明示例性实施例的结合规则来进行机器学习的系统的框图。具体说来，所述系统提出了一种基于融入了规则的预测样本来利用机器学习模型进行预测的处理体系结构，这里，所述规则和机器学习模型涉及相同或相似的预测目标，并且，规则的应用结果被转换为预测样本特征。图1所示的系统可全部通过计算机程序以软件方式来实现，也可由专门的硬件装置来实现，还可通过软硬件结合的方式来实现。相应地，组成图1所示的系统的各个装置可以是仅依靠计算机程序来实现相应功能的虚拟模块，也可以是依靠硬件结构来实现所述功能的通用或专用器件，还可以是运行有相应计算机程序的处理器等。利用所述系统，能够将关于预测目标的规则有效地融入机器学习模型的预测样本中，从而取得较好的预测结果。

[0104] 如图1所示，数据记录获取装置100用于获取数据记录，其中，所述数据记录包括多个属性信息。

[0105] 在本发明的示例性实施例中，在使用机器学习技术针对数据记录的属性信息来得到某种预测结果的过程中，还将结合关于预测目标的规则。

[0106] 作为示例，这里的预测目标可涉及业务判断，例如，预测出是否涉嫌欺诈、信用评分、差异化定价、企业倒闭概率等。举例说来，数据记录可包括个人申请信用卡时填写的个人信息，相应地，预测目标可指示此次申请是否涉及欺诈或指示应批准的信用卡可用额度等。或者，数据记录可包括将采用差异化定价的商品和/或潜在购买者的相关信息，相应地，预测目标可以是所述商品面向所述潜在购买者的个性化预估价格。或者，数据记录可包括企业注册和经营的相关信息，相应地，预测目标可以是该企业在未来预定时间段内倒闭的概率。

[0107] 此外，预测目标也可涉及行为判断，例如，预测出点击概率、营销响应率、推荐响应率等。应注意，这里的预测目标可以是能够预测出的任意内容或事项，也就是说，本发明的示例性实施例在数据记录(及其属性信息)和预测目标方面并不施加任何特定的限制，任何能通过机器学习技术进行预估的数据记录和预测目标均可应用于本发明的示例性实施例。

[0108] 上述数据记录可以是在线产生的数据、预先生成并存储的数据、或者是通过输入装置或传输媒介而从外部接收的数据。这些数据可涉及个人、企业或组织的信息，例如，身份、学历、职业、资产、联系方式、负债、收入、盈利、纳税等信息。或者，这些数据也可涉及业务相关项目的信息，例如，关于买卖合同的交易额、交易双方、标的物、交易地点等信息。

应注意,本发明的示例性实施例中提到的属性信息内容可涉及任何对象或事务在某方面的表现或性质,而不限于对个人、物体、组织、单位、机构、项目、事件等进行限定或描述。

[0109] 数据记录获取装置100可获取不同来源的结构化或非结构化数据,例如,文本数据或数值数据等。获取的数据记录可用于形成机器学习的样例,参与机器学习的训练/预测过程。这些数据可来源于期望获取预测结果的实体内部,例如,来源于期望获取预测结果的银行、企业、学校;这些数据也可来源于上述实体以外,例如,来源于数据提供商、互联网(例如,社交网站)、移动运营商、APP运营商、快递公司、信用机构等。可选地,上述内部数据和外部数据可组合使用,以形成带有更多信息的机器学习样本。

[0110] 上述数据可通过输入装置输入到数据记录获取装置100,或者由数据记录获取装置100根据已有的数据来自动生成,或者可由数据记录获取装置100从网络上(例如,网络上的存储介质(例如,数据仓库))获得,此外,诸如服务器的中间数据交换装置可有助于数据记录获取装置100从外部数据源获取相应的数据。这里,获取的数据可被数据记录获取装置100中的文本分析模块等数据转换模块转换为容易处理的格式。应注意,数据记录获取装置100可被配置为由软件、硬件和/或固件组成的各个模块,这些模块中的某些模块或全部模块可被集成为一体或共同协作以完成特定功能。

[0111] 规则相关特征产生装置200用于将关于预测目标的至少一条规则应用于数据记录的所述多个属性信息,以产生数据记录的规则相关特征。

[0112] 如上所述,根据本发明示例性实施例的所述规则是指与预测目标相关的规则,即,这些规则的规则头与预测目标相同或相关(直接相关或间接相关);相应地,规则相关特征产生装置200旨在针对数据记录的属性信息进行规则检验,例如,判断出数据记录的属性信息是否符合各条规则所限定的规则体内容,并且,规则相关特征产生装置200基于规则检验的结果来产生后续输入机器学习模型的至少一部分样本特征。通过这种方式,能够有效地将规则(例如,反映了专家经验的业务规则)融合于后续的机器学习过程,从而取得较好的预测结果。

[0113] 这里,作为示例,规则相关特征产生装置200可采用适当的方式将数据记录的规则检验结果转换为与该条数据记录相应的机器学习样本的特征,这里的规则检验结果可指示数据记录分别针对各条规则的单独判断结果,也可指示数据记录在多条规则下的综合判断结果,例如,规则检验结果可以是数据记录在所有规则下综合判断出的规则预估结果。相应地,规则相关特征产生装置200可基于上述规则检验结果来产生机器学习样本的至少一部分特征(即,规则相关特征),这些规则相关特征可作为机器学习样本的全部特征,或者,这些规则相关特征还可与其他特征(例如,基于数据记录的属性信息所产生的属性特征)共同构成机器学习样本的特征。

[0114] 具体说来,规则相关产生装置200可获得数据记录在所述规则下的检验结果。这里,在规则涉及数据记录的属性信息自身的情况下,规则相关产生装置200可直接将规则应用于数据记录的属性信息;此外,如果规则涉及的是属性信息(即,属性字段)的变换结果(例如,离散化、字段组合、提取部分字段值、取整等变换结果),则规则相关产生装置200需首先对数据记录的属性信息执行与各条规则相应的变换,并将规则应用于变换后的属性信息。

[0115] 如上所述,规则检验结果可以是每条规则的检验结果,也可以是多条甚至全部规

则的综合检验结果,其中,基于全部规则的综合检验结果可被看做数据记录的规则预测结果。作为示例,规则相关产生装置200可基于马尔可夫逻辑网络获得数据记录基于规则整体所得到的预测结果。

[0116] 上述规则检验结果可作为数据记录的规则相关特征,作为示例,规则相关特征可包括规则预测特征和/或规则描述特征,也就是说,规则相关特征可仅为规则预测特征或规则描述特征,也可以是这两种特征的组合。其中,规则相关产生装置200可基于所述数据记录依据所述至少一条规则而获得的规则预测结果来产生规则预测特征;此外,规则相关产生装置200可基于所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立来产生规则描述特征。

[0117] 作为可选方式,规则相关特征产生装置200在应用所述规则来产生规则相关特征时,可进一步考虑到规则各自的置信度差异。实践中,规则可能是基于先前经验所总结出的关于预测目标的一些规律,这些规律的适用性受到限制,例如,在针对银行客户进行电话营销的案例中,过往的营销规则往往数量有限,无法覆盖到所有客户的具体情况,这也是应用机器学习来提高营销精准性的重要缘由。

[0118] 作为示例,根据本发明的示例性实施例,规则相关特征产生装置200可应用设置了相应权重的各条规则,其中,所述权重表示相应规则的确定性(例如,置信度)。也就是说,规则相关特征产生装置200可将设置了权重的所述至少一条规则应用于所述多个属性信息,使得规则检验结果能够反映出对应规则的置信度。例如,规则相关特征产生装置200可通过将指示所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值分别乘以与所述规则描述特征相应的规则的权重来产生规则描述特征。通过这种方式,可对规则限制进行一定程度上的放松,避免机器学习模型极度依赖规则的可靠性,克服硬性判断失误的弊端,从而得到整体看来更为准确的规则应用结果。

[0119] 这里,所述权重可通过人为指定和/或通过规则学习引擎来设置,其中,所述规则学习引擎用于基于规则训练样本来学习所述至少一条规则的权重。这里,规则训练样本是指已经具有预测目标的实际值的历史数据样本,其可被规则学习引擎用来不断地学习出各条规则的权重。此外,所述规则训练样本作为历史真实样例,也可用作人为指定各条规则权重的依据。

[0120] 作为示例,这里的规则学习引擎可基于马尔可夫逻辑网络,这里的马尔可夫逻辑网络不仅限于马尔可夫逻辑网络的最初版本,也包括诸如概率软逻辑(probabilistic soft logic)等变型或等同体。然而,应注意:根据本发明示例性实施例的规则学习引擎并不受限于马尔可夫逻辑网络的形式,而是可采用任何能够学习出规则权重的形式。

[0121] 作为示例,可由规则相关特征产生装置200在应用规则之前来完成权重的设置,具体说来,规则相关特征产生装置200可针对所述至少一条规则分别设置相应的权重,并将设置了权重的所述至少一条规则应用于所述多个属性信息。这里,规则相关特征产生装置200可通过各种适当的方式来设置各条规则的权重,例如,可通过规则学习的方式来学习或更新各个规则的权重,也可根据业务人员的手动输入来指定各个规则的权重。

[0122] 例如,规则相关特征产生装置200可通过人为指定和/或通过规则学习引擎来针对所述至少一条规则分别设置相应的权重。

[0123] 作为示例,规则相关特征产生装置200可结合规则训练样本集,通过人为指定和/

或通过规则学习引擎来针对所述至少一条规则分别设置相应的权重。这里，规则训练样本集是指已经具有预测目标的实际值的数据样本(即，规则训练样本)的集合，其中，规则训练样本并不限于被用于通过机器学习来获得各规则的权重，而是可以按照任何方式来使用所述规则训练样本。例如，这些规则训练样本可用于帮助业务人员了解各条规则的决策作用，从而设置相应的权重；此外，这些规则训练样本也可用于对规则进行放松，例如，利用基于马尔可夫逻辑网络的规则学习引擎，结合规则训练样本集来学习各条规则的权重。这种规则训练样本集可预先由规则相关特征产生装置200来获取，作为示例，这些规则训练样本与机器学习预测模型的训练样本可来源于相同的数据记录。此外，还可经由预测样本的机器学习结果来补充新的规则训练样本，即，将数据记录连同相应预测样本的预测结果组成新的规则训练样本。

[0124] 这里，作为示例，图1所示的系统还可包括规则获取装置(未示出)，用于获取所述关于预测目标的至少一条规则。这里，作为示例，规则获取装置可输出用于输入规则的图形用户界面，这里的图形用户界面可包括用于手动输入规则的规则编辑界面和/或用于显示规则头和/或规则体的组成项以供手动配置的选择输入型界面。此外，可选地，选择输入型界面还可包括用于手动设置各个规则权重的组件，以便业务人员可以人为指定各个规则的权重。

[0125] 应注意，上述设置权重的方式仅作为示例，而非用于限制本发明示例性实施例的范围，并且，上述方式可单独或组合使用。

[0126] 以上示出了将规则的权重应用于规则相关特征的取值的示例，而根据本发明的示例性实施例，上述规则权重的应用方式并不受限于此。

[0127] 例如，可将规则的权重应用于机器学习预测模型的训练过程，通过这种方式，使得机器学习预测模型的训练阶段可有效地借鉴规则的经验，从而更好地学习出机器学习预测模型。

[0128] 具体说来，规则相关特征产生装置200可将指示所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值作为规则描述特征，相应地，所述至少一条规则之中的各条规则的权重分别作为机器学习预测模型的规则描述特征系数的初始值，这里，所述权重表示相应规则的确定性。

[0129] 此外，在上述示例中，所述至少一条规则之中的各条规则的权重可基于机器学习预测模型的规则描述特征系数的更新值来设置。这里，作为示例，在规则描述特征系数的初始值被设置为相应规则的权重的情况下，在机器学习预测模型的训练过程中，上述系数的数值会不断更新，进而转换为上述规则在机器学习下重新确定的置信度，而该置信度可反向作为规则的权重。

[0130] 具体说来，规则相关特征产生装置200可基于机器学习预测模型的规则描述特征系数的更新值来设置所述至少一条规则之中的各条规则的权重，其中，所述规则描述特征系数用于通过人为指定和/或通过规则学习引擎来针对所述至少一条规则分别设置相应的权重。根据本发明的示例性实施例，规则应用结果可转换为机器学习预测模型的至少一部分样本特征，即，规则描述特征，相应地，机器学习预测模型中涉及规则描述特征的系数可用于反向影响规则权重的设置，例如，这些系数可作为马尔可夫逻辑网络中相应输入的规则权重，进而可基于不断输入的系数来迭代出将应用于属性信息或模型的规则权重。这样，

机器学习与规则系统可在权重方面互相影响,从而迭代得到更为优化的模型。

[0131] 另外,应注意,根据本发明的示例性实施例,所述规则的结果并非仅涉及分类(例如,二分类)判断的结果值,还可以涉及量化的结论值。也就是说,作为示例,规则的结果既可以是取值为“0”或“1”的分类结果,也可以是取值为任意实数(正值或负值均可)的分类结果。特别地,在所述结果与预测目标正相关时,所述结论值为正值,并且,在所述结果与预测目标负相关时,所述结论值为负值。同样地,规则结果的应用方式并不受限于以上作为规则相关特征取值的方式,还可以将规则的结果应用于机器学习预测模型的训练过程,通过这种方式,使得机器学习预测模型的训练阶段可有效地借鉴规则的经验,从而更好地学习出机器学习预测模型。

[0132] 具体说来,规则相关特征产生装置200可将指示所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值作为规则描述特征,相应地,所述至少一条规则之中的各条规则的结论值分别作为机器学习预测模型的规则描述特征系数的初始值,其中,所述结论值表示相应规则的条件成立时所导致的结果。

[0133] 应注意,上述应用规则权重和规则结果的方式还可以进行组合,使得在规则相关特征或机器学习模型的相应系数中体现出规则权重和规则结果的共同影响。作为示例,规则相关特征产生装置200可通过将指示所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值分别乘以与所述规则描述特征相应的规则的权重与结论值的乘积来产生规则描述特征。

[0134] 或者,作为另一示例,规则相关特征产生装置200可将指示所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值作为规则描述特征,相应地,所述至少一条规则之中的各条规则的权重与结论值的乘积分别作为机器学习预测模型的规则描述特征系数的初始值。

[0135] 应注意,在将规则权重和/或规则结论值应用于规则描述特征或相应的模型特征系数初始值时,可适当地调整它们的数值范围,以便更好地适用于机器学习模型。

[0136] 此外,作为可选方式,在通过将规则权重和/或规则结论值设置为相应模型特征系数的初始值的基础上,还可进一步调整所述初始值,以更好地体现算法特点或专家经验。具体说来,机器学习预测模型的规则描述特征系数的初始值可进一步被乘以根据机器学习预测模型的算法而设置的系数修正值。

[0137] 预测样本产生装置300用于至少基于所述规则相关特征来形成预测样本。这里,作为示例,预测样本产生装置300可产生预测样本,使其仅涵盖由规则相关产生装置200产生的规则相关特征。或者,预测样本产生装置300可通过将规则相关特征与其他特征进行组合来产生预测样本,这里,所述其他特征可以是基于数据记录的属性信息而产生的属性特征。这里,属性特征可以是属性信息本身,也可以是通过对属性信息(即,属性字段)进行处理(即,特征处理,例如,离散化、字段组合、提取部分字段值、取整等各种特征工程处理)所得到的结果。

[0138] 机器学习预测装置400用于利用机器学习预测模型,基于预测样本来产生关于预测目标的机器学习预测结果,其中,所述机器学习预测模型被训练为针对预测样本来提供相应的机器学习预测结果。

[0139] 具体说来,在获取了融入有数据记录的规则检验结果的预测样本之后,机器学习

预测装置400可使用之前基于机器学习技术所训练出的机器学习模型(即,机器学习预测模型)来提供关于预测目标的机器学习预测结果。这里,所述机器学习预测模型是基于机器学习算法训练得到的,具体说来,可使用大量的历史数据作为训练样本,基于特定的机器学习算法,训练出机器学习预测模型,其中,训练样本的特征与预测样本的特征相同,相应的预测目标实际值作为训练样本的标记(label)。

[0140] 相应地,当新的预测样本到来时,机器学习预测装置400可将预测样本的特征输入到机器学习预测模型,即可得到所述新的预测样本关于预测目标的预测结果。

[0141] 可以看出,根据本发明的示例性实施例,机器学习预测装置400可使用基于任意机器学习算法训练出的预测模型,这是因为,规则应用结果被转换为预测样本中的特征,确保了原机器学习算法的独立性,相应地,图1所示的系统可理解为是一个通用的机器学习系统,不仅有效地引入了规则判断,而且不必改变原机器学习算法。

[0142] 此外,作为示例,机器学习预测装置400还可将数据记录与机器学习预测结果组合为规则训练样本。如上所述,在本发明的示例性实施例中,可结合规则训练样本集来针对各条规则分别设置相应的权重,上述操作可由规则相关特征产生装置200或其他装置来完成。相应地,通过将数据记录连同相应预测样本的预测结果组成新的规则训练样本,可基于机器学习的预测结果对规则权重进行调整,以更好地克服规则本身的局限性。为此,机器学习预测装置400可将机器学习预测结果作为新规则训练样本的标记,该标记与数据记录则可组合为完整的规则训练样本,相应地,机器学习预测装置400可将组合出的规则训练样本提供给规则相关特征产生装置200。

[0143] 此外,机器学习预测装置400还可提取机器学习预测模型的规则描述特征系数,其中,所述规则描述特征系数用于通过人为指定和/或通过规则学习引擎来针对所述至少一条规则分别设置相应的权重。通过这种方式,机器学习预测装置400可提取出通过机器学习方式得到的相关规则的权重,即,规则描述特征系数,并将提取的系数提供给用于设置规则权重的装置(例如,规则相关特征产生装置200或其他装置),以更新之前设置的规则权重。可以看出,机器学习系统和规则体系在规则系数方面互相影响,有助于得到更优的预估结果。

[0144] 应理解,图1所示的上述装置可被分别配置为执行特定功能的软件、硬件、固件或上述项的任意组合。例如,这些装置可对应于专用的集成电路,也可对应于纯粹的软件代码,还可对应于软件与硬件相结合的单元或模块。此外,这些装置所实现的一个或多个功能也可由物理实体设备(例如,处理器、客户端或服务器等)中的组件来统一执行。

[0145] 此外,为了进一步综合规则和机器学习的预测结果,作为可选方式,可在基于融入了规则相关特征的预测样本得到机器学习预测结果的基础上,进一步融合基于规则整体给出的规则预测结果,以得到最终的预测结果。

[0146] 图2示出根据本发明另一示例性实施例的结合规则来进行机器学习的系统的框图。这里,图2所示的系统中的数据记录获取装置100、规则相关特征产生装置200、预测样本产生装置300和机器学习预测装置400可与图1所示的上述装置执行类似的操作,这里将不再赘述。此外,图2所示的系统还包括规则预测装置500和融合装置600。

[0147] 这里,规则预测装置500用于获得所述数据记录依据所述至少一条规则的规则预测结果,相应地,融合装置600用于将机器学习预测结果与规则预测结果进行融合,以得到

与预测样本相应的融合预测结果。

[0148] 具体说来,规则预测装置500可获得数据记录在全体规则下的规则预测结果。作为示例,规则预测装置500可包括规则学习引擎,该规则学习引擎可基于马尔可夫逻辑网络学习到各条规则的权重,相应地,规则预测装置500将设置了权重的规则应用于数据记录,从而得到该数据记录的规则预测结果。

[0149] 可以看出,作为示例,规则相关特征同样可包括规则预测结果,在这种情况下,规则预测装置500可设置在规则相关特征产生装置200之内,或者说,规则相关产生装置200可实现规则预测装置500的功能,使得系统中不必单独设置规则预测装置500。相应地,所获得的规则预测结果不仅作为规则相关特征(具体说来,规则预测特征)融入预测样本,还将与预测样本的机器学习结果进行融合。

[0150] 此外,作为另一示例,规则相关特征可不包括规则预测结果,在这种情况下,需如图2所示来设置单独的规则预测装置500。

[0151] 融合装置600可从机器学习预测装置400接收机器学习预测结果,从规则预测装置500或规则相关特征产生装置200接收规则预测结果,并将上述两个结果进行融合,以得到与预测样本相应的融合预测结果。

[0152] 此外,作为可选方式,上述融合预测结果也可以用于形成新的规则训练样本,具体说来,融合装置600可将数据记录与融合预测结果组合为规则训练样本。

[0153] 以下参照图3来描述根据本发明示例性实施例的结合规则来进行机器学习的方法的流程图。这里,作为示例,图3所示的方法可由图1所示的系统来执行,也可完全通过计算机程序以软件方式实现,还可通过特定配置的计算装置来执行图3所示的方法。为了描述方便,以下假设图3所示的方法由图1所示的系统来执行。

[0154] 如图所示,在步骤S100中,由数据记录获取装置100获取数据记录,其中,所述数据记录包括多个属性信息。

[0155] 这里,作为示例,每条获取的数据记录可对应于将针对其执行关于预测目标的预估的待预测项(例如,事件或对象),相应地,所述数据记录可包括各种反映事件或对象在某方面的表现或性质(即,属性)的属性信息字段。可对这些属性信息字段进行相应的筛选或其他处理。这里,数据记录获取装置100可通过手动、半自动或全自动的方式来采集数据,或对采集的原始数据进行处理,使得处理后的各种属性信息可后续作为机器学习的样本特征。作为示例,数据记录获取装置100可批量地采集数据。

[0156] 这里,数据记录获取装置100可通过输入装置(例如,工作站)接收用户手动输入的待预测数据记录。此外,数据记录获取装置100可通过全自动的方式从数据源系统地取出待预测数据记录,例如,通过以软件、固件、硬件或其组合实现的定时器机制来系统地请求数据源并从响应中得到所请求的数据。所述数据源可包括一个或多个数据库或其他服务器。可经由内部网络和/或外部网络来实现全自动获取数据的方式,其中可包括通过互联网来传送加密的数据。在服务器、数据库、网络等被配置为彼此通信的情况下,可在没有人工干预的情况下自动进行数据采集,但应注意,在这种方式下仍旧可能存在一定的用户输入操作。半自动方式介于手动方式与全自动方式之间。半自动方式与全自动方式的区别在于由用户激活的触发机制代替了定时器机制。在这种情况下,在接收到特定的用户输入的情况下,才产生提取数据的请求。每次获取数据时,优选地,可将捕获的数据存储在非易失性存储器

中。作为示例,可利用数据仓库来存储在获取期间采集的原始数据以及处理后的数据。

[0157] 上述获取的数据记录可来源于相同或不同的数据源,也就是说,每条数据记录也可以是不同数据记录的拼接结果。例如,除了获取客户向银行申请开通信用卡时填写的信息数据记录(其包括收入、学历、职务、资产情况等属性信息字段)之外,作为示例,数据记录获取装置100可获取该客户在该银行的其他数据记录,例如,贷款记录、日常交易数据等,这些获取的数据记录可拼接为完整的数据记录。此外,数据记录获取装置100还可获取来源于其他私有源或公共源的数据,例如,来源于数据提供商的数据、来源于互联网(例如,社交网站)的数据、来源于移动运营商的数据、来源于APP运营商的数据、来源于快递公司的数据、来源于信用机构的数据等等。

[0158] 可选地,数据记录获取装置100可借助硬件集群(诸如Hadoop集群、Spark集群等)对采集到的数据进行存储和/或处理,例如,存储、分类和其他离线操作。此外,数据记录获取装置100也可对采集的数据进行在线的流处理。

[0159] 作为示例,数据记录获取装置100中可包括文本分析模块等数据转换模块,相应地,在步骤S100中,数据记录获取装置100可将文本等非结构化数据转换为更易于使用的结构化数据以进行进一步的处理或引用。基于文本的数据可包括电子邮件、文档、网页、图形、电子数据表、呼叫中心日志、交易报告等。

[0160] 根据本发明的示例性实施例,作为可选方式,数据记录获取装置100可对获取的数据记录进行特征工程处理,即,对数据记录的属性信息字段值进行处理以获取可用于进行规则学习和/或机器学习的样本特征的属性信息。例如,数据记录获取装置100可对接收到的数据记录的原始属性信息字段进行诸如离散化、字段组合、提取部分字段值、取整等各种特征工程的处理,从而将原始的属性信息字段值转换为可作为规则学习特征和/或机器学习特征的属性信息。

[0161] 在现有技术中,规则参与机器学习主要体现在过滤掉不符合规则的数据、利用规则对原始数据进行变换等,这使得规则不能够参与机器学习模型的训练过程,而只是通过前期对数据的预处理或后期对预测结果的修正来影响最终预测结果。

[0162] 另一方面,根据本发明的示例性实施例,能够在保留机器学习模型独立性的情况下,将规则有效地融入机器学习的过程中,这使得可使用适用于预测目标的任意机器学习模型,而不必局限于某种特定的机器学习模型,实现了机器学习系统的通用性。

[0163] 具体说来,在步骤S200中,由规则相关特征产生装置200将关于预测目标的至少一条规则应用于数据记录的所述多个属性信息,以产生数据记录的规则相关特征。

[0164] 作为示例,在步骤S200之前,还可包括获取所述关于预测目标的至少一条规则的步骤。作为示例,可利用可视化的界面从外部接收规则,进一步地,可在所述界面中展示可供用户配置规则或规则权重的组件,以便用户方便地设置或调整规则。

[0165] 这里,所述至少一条规则整体上与预测目标相关,其中,每条规则可直接或间接地涉及预测目标。

[0166] 例如,假设以预测垃圾邮件为目标,则可至少存在如下两条规则:

[0167] 规则3:垃圾邮件←标题包含“促销”;

[0168] 规则4:垃圾邮件←标题包含“投稿”。

[0169] 又例如,假设以预测癌症为目标,则可至少存在如下两条规则:

[0170] 规则5: 瘤症 (u)  $\leftarrow \forall u$  吸烟者(u);

[0171] 规则6: (吸烟者(u)  $\Leftrightarrow$  吸烟者(v))  $\leftarrow \forall u, v$  朋友(u, v)。

[0172] 根据本发明的示例性实施例,可将这样的规则应用于获取的数据记录,从而得到相应的规则相关特征。这里,假设获取的数据记录的集合为X,每条数据记录可具有d个属性信息,即,第i个数据记录 $x_i \in X$ ,可表示为 $x_i = (x_{i,1}; x_{i,2}; \dots; x_{i,d})$ ,其中,i和d为正整数。相应地,规则相关特征可以是规则作用于这些属性信息之中的至少一部分属性信息而获得的检验结果,这里的检验结果可对应于单条或多条规则,作为示例,还可以是对应于全体规则的规则预测结果。

[0173] 作为示例,可通过判断各条规则的条件是否成立来获得规则相关特征,例如,可将指示所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值作为规则相关特征。具体说来,在上述预测垃圾邮件的示例中,针对第i个数据记录 $x_i$ ,可产生对应的规则相关特征( $R_{i,3}; R_{i,4}$ ),其中, $R_{i,3}$ 可指示规则3的条件是否成立, $R_{i,4}$ 可指示规则4的条件是否成立。举例说来,当第i个数据记录 $x_i$ 的标题信息包含“促销”,即,规则3的规则体成立时, $R_{i,3}$ 可取值为1;当所述标题信息不包含“促销”时, $R_{i,3}$ 可取值为0。此外,当第i个数据记录 $x_i$ 的标题信息包含“投稿”,即,规则4的规则体成立时, $R_{i,4}$ 可取值为1;当所述标题信息不包含“投稿”时, $R_{i,4}$ 可取值为0。

[0174] 实践中,规则往往具有不确定性,比如,标题包含“促销”或“投稿”的邮件并非一定是垃圾邮件。规则本身的不确定性容易导致最终的预测结果与客观情况出入很大,并且,持续使用这样的规则会导致预测结果不断恶化。

[0175] 为此,根据本发明的示例性实施例,在步骤S200中将规则应用于数据记录时,可考虑规则的置信度,具体说来,可将设置了权重的所述至少一条规则应用于所述多个属性信息,使得不同规则对应的置信度在后续的机器学习模型中能够得以区分。这里的权重可通过人为指定和/或通过规则学习引擎而预先设置。

[0176] 具体说来,在步骤S200中,可通过将指示所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值分别乘以与所述规则描述特征相应的规则的权重来产生规则描述特征。例如,在上述预测垃圾邮件的示例中,规则3可被设置了权重0.8,而规则4可被设置了权重0.3,相应地,针对数据记录,当规则3的规则体成立时, $R_{i,3}$ 可取值为0.8;当规则3的规则体不成立时, $R_{i,3}$ 可取值为0。此外,当规则4的规则体成立时, $R_{i,4}$ 可取值为0.3;当规则4的规则体不成立时, $R_{i,4}$ 可取值为0。

[0177] 除了将指示规则的条件是否成立的逻辑值本身作为规则相关特征之外,还可将条件成立时导致的结果引入规则相关特征,特别地,这种规则相关特征可有效地体现规则条件成立时导致的量化结论值。具体说来,在步骤S200中,可通过将指示所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值分别乘以与所述规则描述特征相应的规则的结论值来产生规则描述特征。

[0178] 例如,在预测垃圾邮件的另一示例中,可能存在能够判断垃圾邮件概率的规则,例如:

[0179] 规则33:70%为垃圾邮件 $\leftarrow$ 标题包含“促销”;

[0180] 规则44:40%为垃圾邮件 $\leftarrow$ 标题包含“投稿”。

[0181] 相应地,针对第*i*个数据记录 $x_i$ ,可产生对应的规则相关特征( $R_{i,33}; R_{i,44}$ ),其中, $R_{i,33}$ 可对应于规则33的条件成立所导致的结果, $R_{i,44}$ 可对应于规则44的条件成立所导致的结果。举例说来,当第*i*个数据记录 $x_i$ 的标题信息包含“促销”,即,规则33的规则体成立时, $R_{i,33}$ 可取值为0.7;当所述标题信息不包含“促销”时, $R_{i,33}$ 可取值为0。此外,当第*i*个数据记录 $x_i$ 的标题信息包含“投稿”,即,规则44的规则体成立时, $R_{i,44}$ 可取值为0.4;当所述标题信息不包含“投稿”时, $R_{i,44}$ 可取值为0。

[0182] 在如上应用规则的结果值的情况下,还可进一步结合各条规则的权重,具体说来,在步骤S200中,可通过将指示所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值分别乘以与所述规则描述特征相应的规则的权重与结论值的乘积来产生规则描述特征。

[0183] 例如,假设上述规则 $R_{i,33}$ 被设置了权重0.6,而规则 $R_{i,44}$ 被设置了权重0.5。相应地,当第*i*个数据记录 $x_i$ 的标题信息包含“促销”,即,规则33的规则体成立时, $R_{i,33}$ 可取值为 $0.7 * 0.6 = 0.42$ ;当所述标题信息不包含“促销”时, $R_{i,33}$ 可取值为0。此外,当第*i*个数据记录 $x_i$ 的标题信息包含“投稿”,即,规则44的规则体成立时, $R_{i,44}$ 可取值为 $0.4 * 0.5 = 0.2$ ;当所述标题信息不包含“投稿”时, $R_{i,44}$ 可取值为0。

[0184] 应注意,在确定规则相关特征的具体特征值时,可适当地调整权重或结论值的取值范围,使得特征值能够有效地适用于机器学习运算。

[0185] 实质上,上述规则相关特征可表征为一种规则描述特征,该规则描述特征可基于所述数据记录针对所述至少一条规则之中的各条规则的条件是否成立来产生,可指示规则的条件是否成立的结果,也可进一步结合规则的权重和/或结论值。

[0186] 可以看出,在上述示例中,针对数据记录,每条规则具有相应的规则描述特征,该规则描述特征的取值指示所述规则应用于数据记录的属性信息时的检验结果。然而,本发明的示例性实施例并非受限于此,规则描述特征可对应于多条规则被应用后的综合结果。应注意,规则相关特征的产生方式并不受限于上述示例,任何通过将规则应用于数据记录而产生相关特征的方式均可应用于本发明的示例性实施例。

[0187] 例如,规则相关特征还可表征为一种规则预测特征,该规则预测特征基于所述数据记录依据所述至少一条规则而获得的规则预测结果来产生。例如,在系统中(例如,在规则相关特征产生装置200内)可实施有规则学习引擎,通过该规则学习引擎,可在步骤S200中获得数据记录基于规则整体而得到的预测结果,该预测结果可作为规则相关特征。

[0188] 例如,在上述预测垃圾邮件的示例中,在步骤S200中,可利用规则学习引擎,基于包括规则3、规则4以及其他相关规则的整体规则(可选地,连同相应的权重),预测出某封邮件为垃圾邮件的概率为0.4,并将该预测结果0.4作为所述邮件的规则相关特征PRUL。应注意,规则预测特征可连同规则描述特征一起作为数据记录的规则相关特征。

[0189] 根据本发明的示例性实施例,规则学习引擎可用于基于输入的规则和相应的实例(即,作为规则训练样本的真实历史数据记录),学习得到每个规则的权重,并能够基于赋予了权重的规则来针对新的数据记录给出预测结果(例如,预估概率)。这里,输入的规则可包含权重,也可不包含权重。这里的规则权重可初始由人为指定,并且,随后利用诸如马尔可夫逻辑网络等对人为指定的规则进行不断更新。此外,这里的规则权重还可来源于后续的机器学习模型,即,将训练过程期间或训练完成之后得到的机器学习模型中的规则相关特

征(例如,规则描述特征)的系数反馈给规则学习引擎,从而使得规则学习部分和机器学习部分通过规则特征的权重而相互交互,迭代出性能更优的模型。

[0190] 作为示例,上述规则学习引擎可基于马尔可夫逻辑网络(或其变型,例如,概率软逻辑)而获得,也可以按照其他形式来构造所述规则学习引擎。具体说来,马尔可夫逻辑网络是由一阶逻辑公式的规则及其对应的权重值组成的二元组集合。马尔可夫逻辑网络的基本思想是将一阶逻辑公式的限制放松,即,一个事件违反的公式越多,其发生概率越小,但未必完全不会发生(即,发生概率未必为0)。通过基于规则训练样本集将所述规则进行实例化,可得到马尔可夫逻辑网络,进而能够在马尔可夫逻辑网络上进行学习和推理。

[0191] 图4示出根据本发明示例性实施例的马尔可夫逻辑网络的示例。在图4所示的示例中,以预测癌症为目标,相应地,存在如下两条规则:

[0192] 规则5: 瘤症 (u)  $\leftarrow \forall u$  吸烟者(u);

[0193] 规则6: ( $\text{吸烟者}(u) \Leftrightarrow \text{吸烟者}(v)$ )  $\leftarrow \forall u, v$  朋友(u, v)。

[0194] 在图4中示出了A、B两个人在上述规则下的实例化情况,本领域技术人员应知晓,在一组涉及上述规则的实例组成的世界,可学习出各条规则对应的权重以及基于规则整体所预测出的预估结果。

[0195] 可以看出,根据本发明的示例性实施例,可将马尔可夫逻辑网络学习到的规则权重及其预测结果应用于一般的机器学习问题(例如,分类问题等),从而与通用的机器学习方法相结合,不再整体依赖于马尔可夫随机场,也不仅限于解决统计关系学习及推理问题。

[0196] 再次参照图3,在步骤S300中,由预测样本产生装置300至少基于所述规则相关特征来形成预测样本。通过这种方式,可扩展原始的数据记录以容纳与规则相关的样本特征,从而使得规则能够直接有效地参与到模型的预测过程。

[0197] 这里,预测样本至少涵盖规则相关特征,此外,预测样本还可包括其他的特征,例如,基于属性信息而获得的属性特征。作为示例,预测样本产生装置300可通过将规则相关特征与其他属性特征串联来产生预测样本。

[0198] 例如,假设数据记录 $x_i = (x_{i,1}; x_{i,2}; \dots; x_{i,d})$ ,其经过m(其中,m为正整数)条规则的检验,可分别得到m个规则相关特征 $(r_{i,1}; r_{i,2}; \dots; r_{i,m})$ ,其中,每一个规则相关特征的取值取决于相应规则应用于数据记录时的检验结果,例如,当相应规则的规则体成立时,规则相关特征可取值为1、规则的权重值、规则的结论值、规则的权重值与结论值的乘积等;否则,规则相关特征可取值为0。相应地,预测样本产生装置300可产生基于数据记录的预测样本 $y_i = (x_{i,1}; x_{i,2}; \dots; x_{i,d}; r_{i,1}; r_{i,2}; \dots; r_{i,m})$ 。

[0199] 又例如,假设数据记录 $x_i = (x_{i,1}; x_{i,2}; \dots; x_{i,d})$ ,其经过m条规则的综合检验,可得到关于预测目标的预测结果,该预测结果可作为规则相关特征 $P_{RULi}$ 。相应地,预测样本产生装置300可产生基于数据记录的预测样本 $y_i = (x_{i,1}; x_{i,2}; \dots; x_{i,d}; P_{RULi})$ 。

[0200] 或者,上述规则描述特征和规则预测特征两者可同时作为规则相关特征,相应地,预测样本产生装置300可产生基于数据记录 $x_i$ 的预测样本 $y_i = (x_{i,1}; x_{i,2}; \dots; x_{i,d}; r_{i,1}; r_{i,2}; \dots; r_{i,m}; P_{RULi})$ 。

[0201] 应注意,预测样本产生装置300在产生预测样本时,并不限于将规则相关特征与其他特征进行串联,还可以采用各种适当的特征处理方式(例如,特征组合等)。

[0202] 接下来,在步骤S400中,由机器学习预测装置400利用机器学习预测模型,基于预测样本来产生关于预测目标的机器学习预测结果,其中,所述机器学习预测模型被训练为针对预测样本来提供相应的机器学习预测结果。

[0203] 具体说来,机器学习预测装置400可将预测样本输入机器学习预测模型,以得到关于预测目标的机器学习预测结果。这里所述的机器学习预测模型可以是适用于原始数据记录的任意机器学习模型,例如,如果原始数据记录适用于支持向量机(Support Vector Machine)、对数几率回归(Logistic Regression)等模型,则机器学习预测模型也可以采用相同的算法,而不会因为规则相关特征的引入而有所限制。

[0204] 作为示例,机器学习预测模型可以是预先训练出的预测模型。具体说来,对于历史上得到过关于预测目标的真实结果的数据记录(例如,最终确定的垃圾邮件等),将这些真实结果作为有监督学习下的标记(label),并将对应的预测样本特征作为相应的训练样本特征,进而可训练出用于结合了规则相关特征的预测样本来进行预测的机器学习预测模型。

[0205] 这里,机器学习预测模型可以由机器学习预测装置400预先训练得出。此外,机器学习预测模型也可以由设置在图1所示的系统中的与模型训练相关的装置(未示出)预先训练得出,或可以由图1所示的系统以外的外部装置预先训练得出,在这样的情况下,机器学习预测装置400可从模型训练相关装置或外部装置接收其所训练出的机器学习预测模型。

[0206] 在步骤S400中获得关于预测目标的预测结果之后,机器学习预测装置400可将机器学习预测结果存储在相应的存储器中以便在后续进行进一步的处理,或者,这些机器学习预测结果可被发送到外部的处理装置。此外,也可通过输出装置将机器学习预测结果展示给用户。

[0207] 根据本发明的示例性实施例,可将机器学习的中间或最终成果反馈给规则部分。作为示例,所述方法可还包括以下步骤:将数据记录与机器学习预测结果组合为规则训练样本。作为另一示例,所述方法可还包括以下步骤:提取机器学习预测模型的规则描述特征系数的更新值,用于设置所述至少一条规则之中的各条规则的权重。

[0208] 可以看出,在本发明的示例性实施例中,机器学习预测结果可被视作能够更新规则权重的规则训练样本(即,实例)的标记,并且/或者,机器学习模型本身的规则相关特征系数也可被反馈到规则系统,以直接作用于权重的更新。通过这种方式,机器学习系统和规则系统可互相影响,并反复迭代出效果更好的模型。

[0209] 根据本发明的示例性实施例,除了在预测样本的特征层面融合有规则之外,还可在预测结果的层面将机器学习和规则进一步融合。也就是说,可将机器学习预测结果和规则预测结果进行融合,以代替机器学习预测结果来作为最终的预测结果。

[0210] 图5示出根据本发明另一示例性实施例的结合规则来进行机器学习的方法的流程图。参照图5,步骤S100到步骤S400与图3所示的步骤S100到步骤S400基本类似,在这里将不再赘述。

[0211] 图5所示的方法还包括步骤S600,在该步骤中,由融合装置600将机器学习预测结果与规则预测结果进行融合,以得到与预测样本相应的融合预测结果。这里,机器学习预测结果可来源于步骤S400,而规则预测结果可来源于步骤S200(在规则相关特征包括规则预测结果的情况下)或其他步骤。也就是说,在规则相关特征不包括规则预测结果的情况下,

图5所示的方法还包括以下步骤：依据所述至少一条规则来获得所述数据记录的规则预测结果。相应地，可在步骤S600中将机器学习预测结果与规则预测结果融合为最终的预测结果。作为示例，假设针对数据记录 $x_i$ ，规则预测结果为 $P_{RULi}$ 且机器学习预测结果为 $P_{MLi}$ ，则在步骤S600中，融合装置600可通过诸如加权平均等方式将上述结果融合为 $P=w \times P_{RULi} + (1-w) \times P_{MLi}$ ，其中， $0 \leq w \leq 1$ 。这里，可根据需要来设置w的具体取值。例如，假设针对某条数据记录，规则预测结果为0.4，机器学习预测结果为0.7，如果融合方式为取平均值，则融合结果为0.55。

[0212] 应注意，上述机器学习预测结果不仅可指示特征层面融合了规则的情况下所得到的预测结果，还可包括不涉及规则的单纯机器学习预测结果。

[0213] 此外，根据本发明的示例性实施例，可将融合结果反馈给规则处理部分，以用于更新规则的权重。作为示例，图5所示的方法可还包括以下步骤：将数据记录与融合结果组合为规则训练样本。

[0214] 应注意，根据本发明示例性实施例的上述结合规则来进行机器学习的系统可完全依赖计算机程序的运行来实现相应的功能，即，各个装置与计算机程序的功能架构中与各步骤相应，使得整个系统通过专门的软件包（例如，lib库）而被调用，以实现相应的预测功能。

[0215] 图6示出本发明另一示例性实施例的结合规则来进行机器学习的系统的框图。这里，本领域技术人员应理解，图6所示的系统旨在训练出根据本发明示例性实施例的机器学习预测模型，其所实现的处理可对应于图1所示的系统所执行的处理。例如，构成上述系统的装置以及装置所执行的操作可具有对应性，相应地，对于图6中各装置的具体操作，以下将参照图1所示的系统来进行描述，而不再重复相关的细节。

[0216] 具体说来，图6所示的系统包括：历史数据记录获取装置1000、规则相关特征产生装置2000、训练样本产生装置3000和机器学习模型训练装置4000。

[0217] 历史数据记录获取装置1000用于获取历史数据记录，其中，所述历史数据记录包括多个属性信息和作为预测目标实际值的标记。这里，应理解，历史数据记录获取装置1000可按照类似于数据记录获取装置100的方式进行操作，只是其所获取的是已经具有了预测目标实际值的历史数据。

[0218] 规则相关特征产生装置2000用于将关于预测目标的至少一条规则应用于所述多个属性信息，以产生历史数据记录的规则相关特征。这里，应理解，规则相关特征产生装置2000旨在产生模型训练样本中的规则相关特征，具体操作可对应于图1的规则相关特征产生装置200产生模型预测样本中的规则相关特征的方式。

[0219] 训练样本产生装置3000用于至少基于所述规则相关特征和标记来形成训练样本。这里，在形式上，训练样本与预测样本相比，还包括历史数据记录中的标记，相应地，应理解，训练样本产生装置3000可按照与预测样本产生装置300类似的方式进行操作，只是还需在训练样本中包含标记信息。

[0220] 机器学习模型训练装置4000用于基于训练样本来训练机器学习预测模型，其中，所述机器学习模型用于针对新的数据记录来提供关于预测目标的机器学习预测结果。

[0221] 如上所述，历史数据记录获取装置1000、规则相关特征产生装置2000、训练样本产生装置3000和机器学习模型训练装置4000在具体操作上可与图1所示的数据记录获取装置

100、规则相关特征产生装置200、预测样本产生装置300和机器学习预测装置400相类似,使得本领域技术人员能够获知相应的处理细节。此外,图6所示的系统也可如图1所示的系统一样,还包括诸如规则获取装置的额外装置,其中,所述规则获取装置用于获取所述关于预测目标的至少一条规则。

[0222] 如上所述,历史数据记录获取装置1000、规则相关特征产生装置2000、训练样本产生装置3000和机器学习模型训练装置4000可分别与图1所示的数据记录获取装置100、规则相关特征产生装置200、预测样本产生装置300和机器学习预测装置400相集成,使得集成后的装置在模型训练和模型预测阶段分别执行相应的操作。此外,图6所示的系统中的至少一个装置或全部装置也可与图1所示的系统相互分离,作为独立的部分存在。

[0223] 根据本发明的示例性实施例,在机器学习模型的训练过程中,使用了包含规则相关特征的训练样本,即,规则转换为特征,从而更为有效地参与到机器学习过程中。

[0224] 以下将参照图7来描述根据本发明另一示例性实施例的结合规则来进行机器学习的方法的流程图。这里,作为示例,图7所示的方法可由图6所示的系统来执行,也可完全通过计算机程序以软件方式实现,还可通过特定配置的计算装置来执行图7所示的方法。为了描述方便,以下假设图7所示的方法由图1所示的系统来执行。

[0225] 这里,本领域技术人员应理解,图7所示的方法旨在训练出根据本发明示例性实施例的机器学习预测模型,其所实现的处理可对应于图3所示的方法中所执行的处理。相应地,对于图7中各步骤的具体操作,以下将参照图3所示的方法来进行描述,而不再重复相关的细节,仅重点描述图3中未体现出的技术特征。

[0226] 如图所示,在步骤S1000中,由历史数据记录获取装置1000获取历史数据记录,其中,所述历史数据记录包括多个属性信息和作为预测目标实际值的标记。这里,应理解,在步骤S1000中,历史数据记录获取装置1000可按照类似于数据记录获取装置100在步骤S100中执行的操作来获取历史数据记录,只是其所获取的是已经具有了预测目标实际值的历史数据。

[0227] 接下来,在步骤S2000中,由规则相关特征产生装置2000将关于预测目标的至少一条规则应用于所述多个属性信息,以产生历史数据记录的规则相关特征。这里,在步骤S2000中,规则相关特征产生装置2000可按照类似于规则相关特征产生装置200在步骤S200中执行的操作来产生同样的规则相关特征。

[0228] 在步骤S3000中,由训练样本产生装置3000至少基于所述规则相关特征和标记来形成训练样本。这里,在步骤S3000中,训练样本产生装置3000可按照类似于预测样本产生装置300在步骤S300中执行的操作来产生训练样本,只是所述训练样本还需要包括相应的标记。

[0229] 在步骤S4000中,由机器学习模型训练装置4000基于训练样本来训练机器学习预测模型,其中,所述机器学习模型用于针对新的数据记录来提供关于预测目标的机器学习预测结果。

[0230] 这里,在训练机器学习预测模型时,作为可选方式,可进一步将规则应用于模型的系数,例如,将规则的权重和/或规则的结论值应用于模型的系数。通过这种方式,可有效地借鉴先验知识,有利于快速准确地学习出机器学习模型。

[0231] 具体说来,对于训练样本中的规则描述特征而言,其在模型中的对应系数的初始

值可被设置为与规则相关。

[0232] 作为示例,在步骤S2000中,规则相关特征产生装置2000可将指示历史数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值作为规则描述特征,相应地,在步骤S4000中,机器学习模型训练装置4000可将所述至少一条规则之中的各条规则的权重分别作为机器学习预测模型的规则描述特征系数的初始值,其中,所述权重表示相应规则的确定性。

[0233] 在上述示例中,还可基于机器学习预测模型的规则描述特征系数的更新值来设置所述至少一条规则之中的各条规则,从而使得机器学习与规则两者能够互相影响,从而迭代出更好的模型。

[0234] 作为另一示例,在步骤S2000中,规则相关特征产生装置2000可将指示所述历史数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值作为规则描述特征,相应地,在步骤S4000中,机器学习模型训练装置4000可将所述至少一条规则之中的各条规则的结论值分别作为机器学习预测模型的规则描述特征系数的初始值,其中,所述结论值表示相应规则的条件成立时所导致的结果。

[0235] 作为又一示例,在步骤S2000中,规则相关特征产生装置2000可将指示所述历史数据记录针对所述至少一条规则之中的各条规则的条件是否成立的逻辑值作为规则描述特征,相应地,在步骤S4000中,机器学习模型训练装置4000可将所述至少一条规则之中的各条规则的权重与结论值的乘积分别作为机器学习预测模型的规则描述特征系数的初始值。

[0236] 除此之外,还可通过系数修正值来调整规则描述特征系数的初始值。具体说来,在训练机器学习预测模型时,所述方法还可包括以下步骤:将机器学习预测模型的规则描述特征系数的初始值乘以系数修正值,从而使得模型的训练过程更为有效。这里,系数修正值可由编程人员手动调节,或者,也可根据机器学习预测模型的算法而自动设置。

[0237] 另一方面,图1、图2或图6所示的各个装置也可以通过硬件、软件、固件、中间件、微代码或其任意组合来实现。当以软件、固件、中间件或微代码实现时,用于执行相应操作的程序代码或者代码段可以存储在诸如存储介质的计算机可读介质中,使得处理器可通过读取并运行相应的程序代码或者代码段来执行相应的操作。

[0238] 这里,本发明的示例性实施例还可以实现为计算装置,该计算装置包括存储部件和处理器,存储部件中存储有计算机可执行指令集合,当所述计算机可执行指令集合被所述处理器执行时,执行上述结合规则来进行机器学习的方法。

[0239] 具体说来,所述计算装置可以部署在服务器或客户端中,也可以部署在分布式网络环境中的节点装置上。此外,所述计算装置可以是PC计算机、平板装置、个人数字助理、智能手机或其他能够执行上述指令集合的装置。

[0240] 这里,所述计算装置并非必须是单个的计算装置,还可以是任何能够单独或联合执行上述指令(或指令集)的装置或电路的集合体。计算装置还可以是集成控制系统或系统管理器的一部分,或者可被配置为与本地或远程(例如,经由无线传输)以接口互联的便携式电子装置。

[0241] 在所述计算装置中,处理器可包括中央处理器(CPU)、图形处理器(GPU)、可编程逻辑装置、专用处理器系统、微控制器或微处理器。作为示例而非限制,处理器还可包括模拟处理器、数字处理器、微处理器、多核处理器、处理器阵列、网络处理器等。

[0242] 上述结合规则来进行机器学习的方法中所描述的某些操作可通过软件方式来实现,某些操作可通过硬件方式来实现,此外,还可通过软硬件结合的方式来实现这些操作。

[0243] 处理器可运行存储在存储部件之一中的指令或代码,其中,所述存储部件还可以存储数据。指令和数据还可经由网络接口装置而通过网络被发送和接收,其中,所述网络接口装置可采用任何已知的传输协议。

[0244] 存储部件可与处理器集成为一体,例如,将RAM或闪存布置在集成电路微处理器等之内。此外,存储部件可包括独立的装置,诸如,外部盘驱动、存储阵列或任何数据库系统可使用的其他存储装置。存储部件和处理器可在操作上进行耦合,或者可例如通过I/O端口、网络连接等互相通信,使得处理器能够读取存储在存储部件中的文件。

[0245] 此外,所述计算装置还可包括视频显示器(诸如,液晶显示器)和用户交互接口(诸如,键盘、鼠标、触摸输入装置等)。计算装置的所有组件可经由总线和/或网络而彼此连接。

[0246] 上述关于结合规则来进行机器学习的方法所涉及的操作可被描述为各种互联或耦合的功能块或功能示图。然而,这些功能块或功能示图可被均等地集成为单个的逻辑装置或按照非确切的边界进行操作。

[0247] 具体说来,如上所述,根据本发明示例性实施例的结合规则来进行机器学习的计算装置可包括存储部件和处理器,存储部件中存储有计算机可执行指令集合,当所述计算机可执行指令集合被所述处理器执行时,执行下述步骤:(A)获取数据记录,其中,所述数据记录包括多个属性信息;(B)将关于预测目标的至少一条规则应用于所述多个属性信息,以产生数据记录的规则相关特征;(C)至少基于所述规则相关特征来形成预测样本;以及(D)利用机器学习预测模型,基于预测样本来产生关于预测目标的机器学习预测结果,其中,所述机器学习预测模型被训练为针对预测样本来提供相应的机器学习预测结果。应注意,以上已经结合图3到图5描述了根据本发明示例性实施例的结合规则来进行机器学习的方法的各处理细节,这里将不再赘述计算装置执行各步骤时的处理细节。

[0248] 此外,根据本发明另一示例性实施例的结合规则来进行机器学习的计算装置可包括存储部件和处理器,存储部件中存储有计算机可执行指令集合,当所述计算机可执行指令集合被所述处理器执行时,执行下述步骤:(A)获取历史数据记录,其中,所述历史数据记录包括多个属性信息和作为预测目标实际值的标记;(B)将关于预测目标的至少一条规则应用于所述多个属性信息,以产生历史数据记录的规则相关特征;(C)至少基于所述规则相关特征和标记来形成训练样本;以及(D)基于训练样本来训练机器学习预测模型,其中,所述机器学习模型用于针对新的数据记录来提供关于预测目标的机器学习预测结果。

[0249] 以上已经描述了本发明的各示例性实施例,应理解,上述描述仅是示例性的,并非穷尽性的,并且本发明也不限于所披露的各示例性实施例。在不偏离本发明的范围和精神的情况下,对于本领域技术人员来说许多修改和变更都是显而易见的。因此,本发明的保护范围应该以权利要求的范围为准。

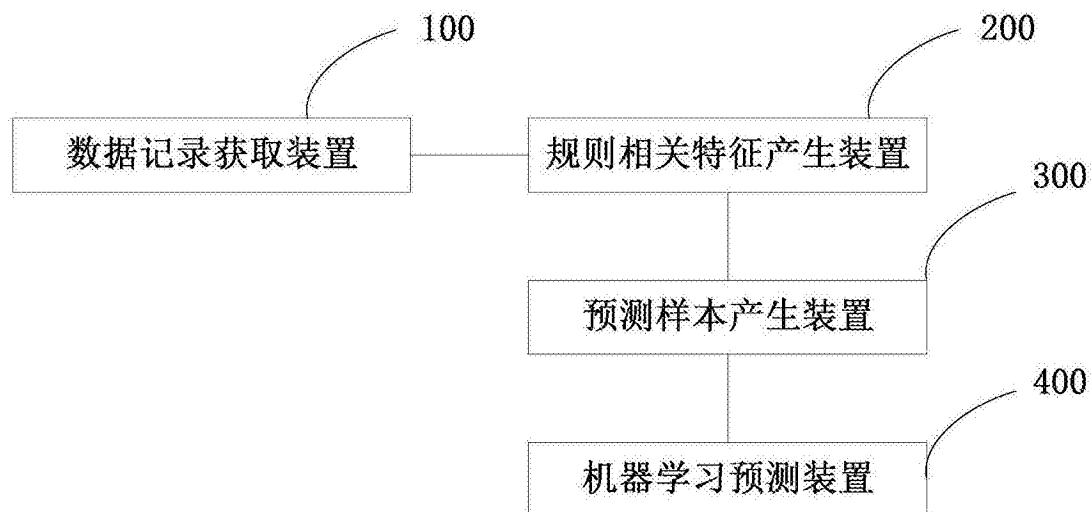


图1

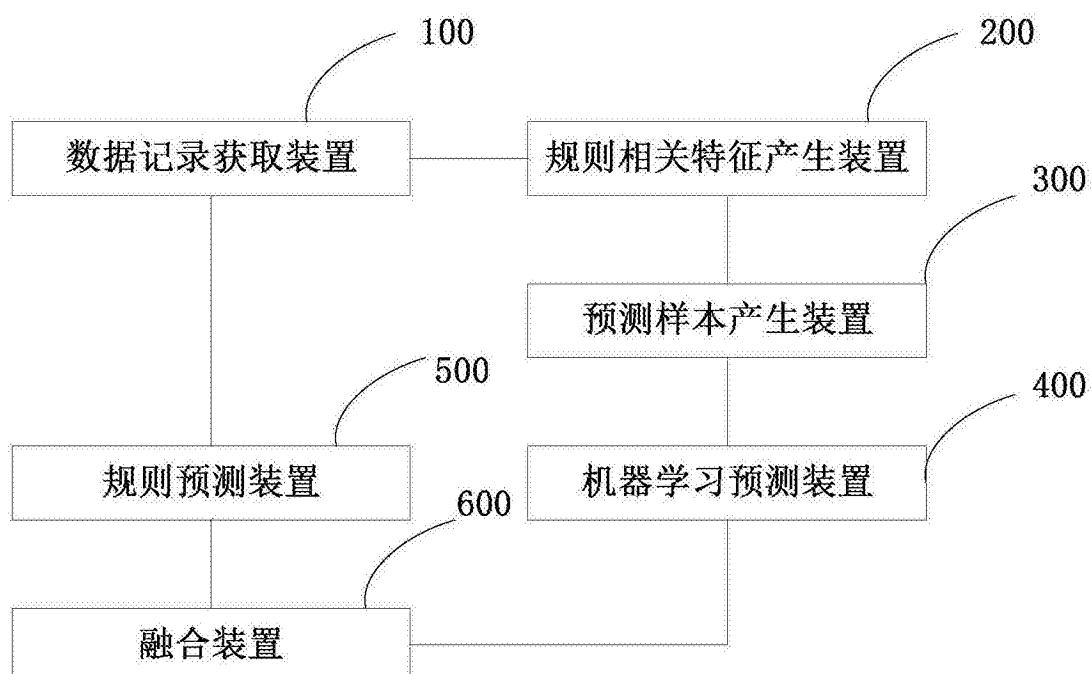


图2

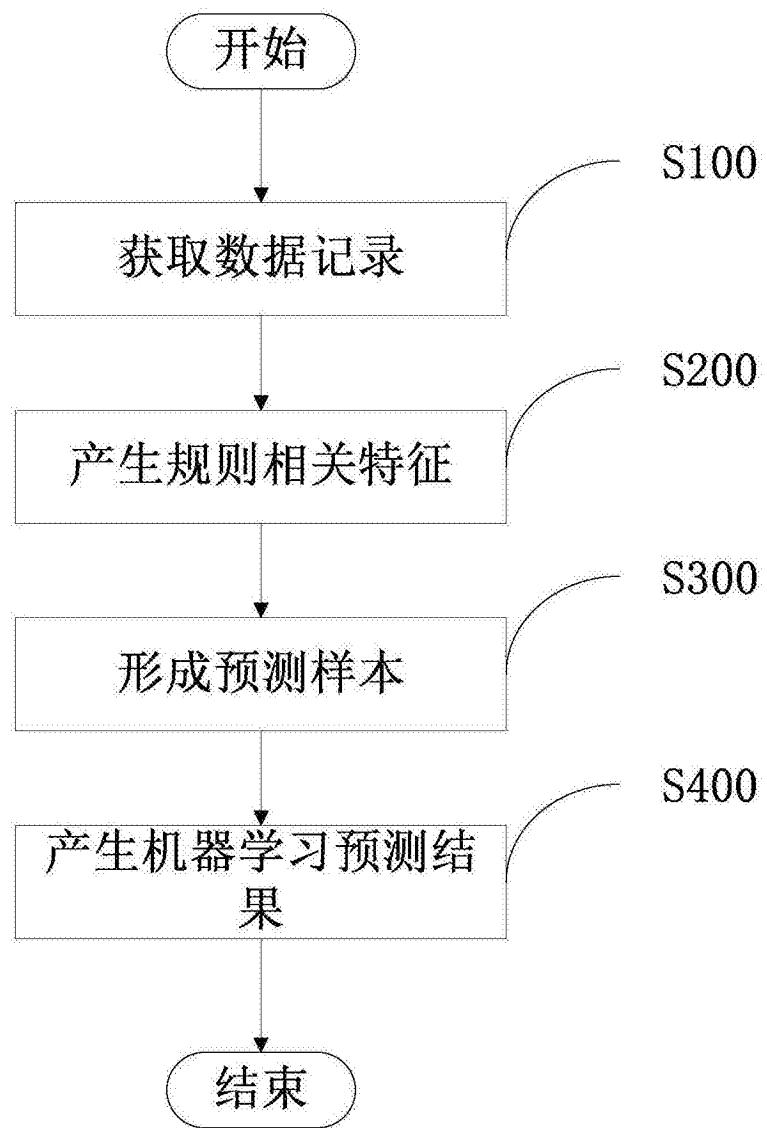


图3

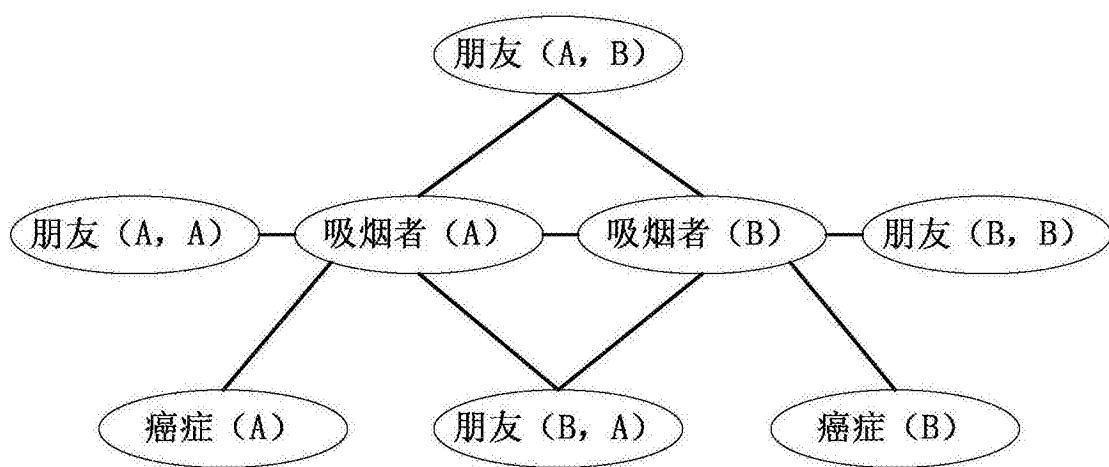


图4

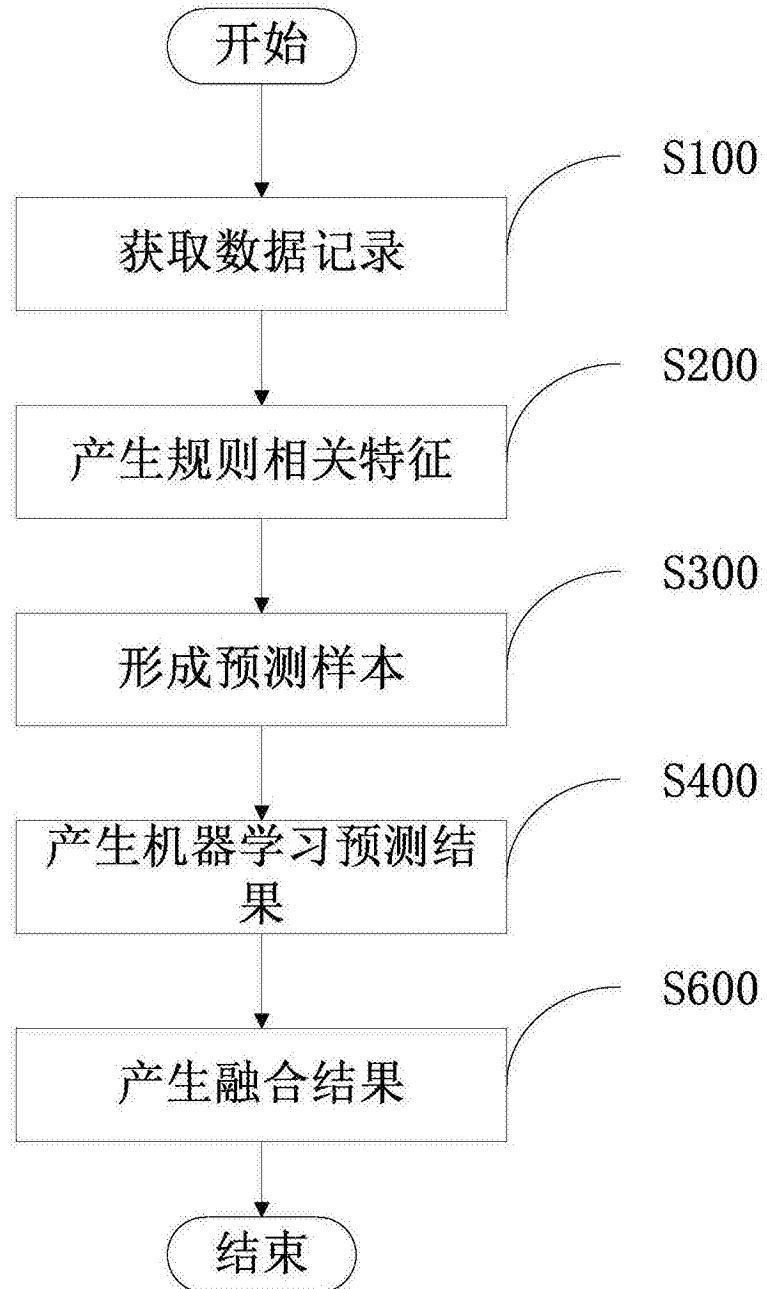


图5

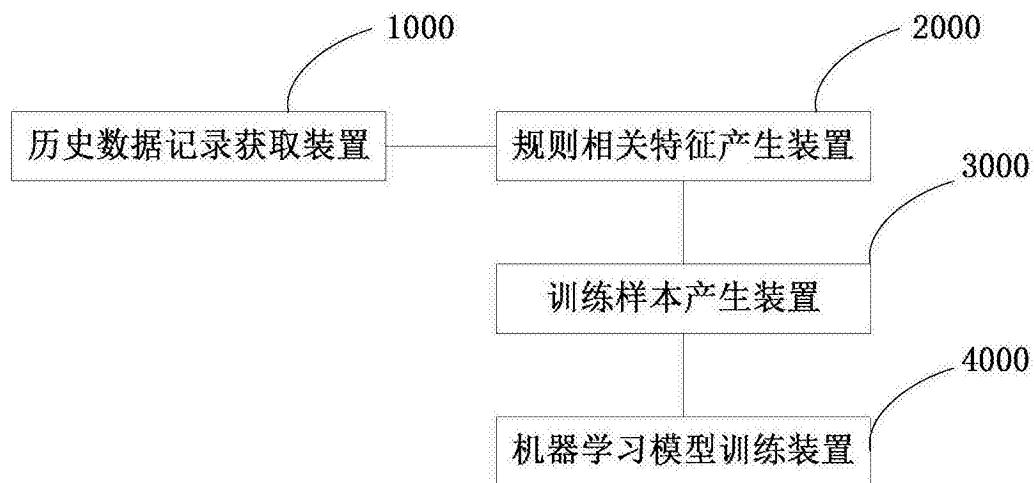


图6

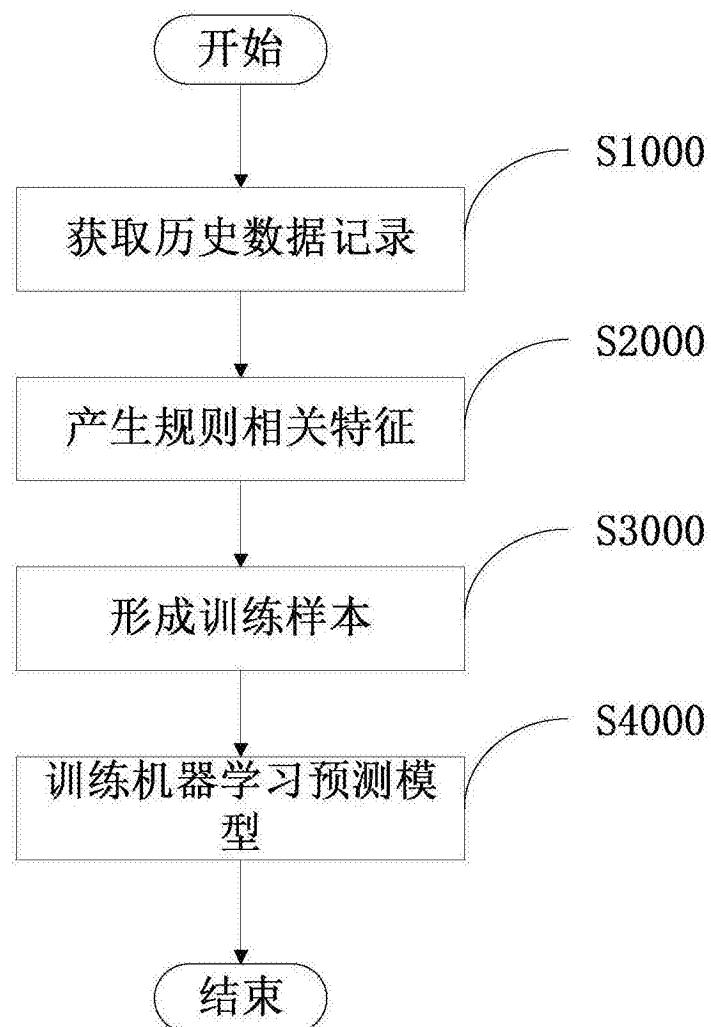


图7