



(12)发明专利申请

(10)申请公布号 CN 111274472 A

(43)申请公布日 2020.06.12

(21)申请号 201811475043.X

(22)申请日 2018.12.04

(71)申请人 北京嘀嘀无限科技发展有限公司
地址 100193 北京市海淀区东北旺西路8号
院34号楼

(72)发明人 付俊强 杜龙志 王志明 李樊
卓呈祥 郗小虎

(74)专利代理机构 北京超成律师事务所 11646
代理人 邓超

(51)Int.Cl.
G06F 16/9535(2019.01)
G06K 9/62(2006.01)

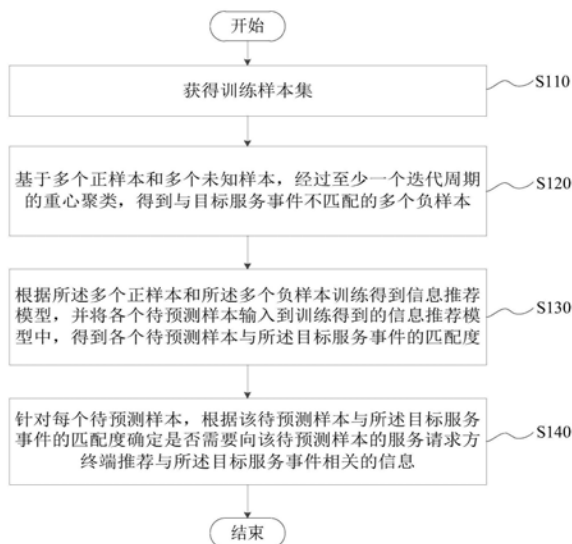
权利要求书7页 说明书19页 附图6页

(54)发明名称

信息推荐方法、装置、服务器及可读存储介质

(57)摘要

本申请实施例提供了一种信息推荐方法、装置、服务器及可读存储介质,通过对训练样本集中的多个正样本和多个未知样本经过至少一个迭代周期的重心聚类,可以得到更为准确的多个与目标服务事件不匹配的负样本,避免在训练信息推荐模型过程中错误地将原本与目标服务事件匹配的样本作为负样本的特征信息进行学习而导致的预测误差。然后按照以此训练得到的信息推荐模型预测各个待预测样本与目标服务事件的匹配度,从而确定是否需要向各个待预测样本的服务请求方终端推荐与目标服务事件相关的信息,如此,能够提高信息推荐的精准性,同时为服务请求方提供更多便利、高效的目标服务。



1. 一种信息推荐方法,其特征在于,应用于服务器,所述方法包括:
 - 获得训练样本集,所述训练样本集包括多个正样本和多个未知样本;
 - 基于所述多个正样本和多个未知样本,经过至少一个迭代周期的重心聚类,得到与目标服务事件不匹配的多个负样本,其中,每个正样本与所述目标服务事件匹配;
 - 根据所述多个正样本和所述多个负样本训练得到信息推荐模型,并将各个待预测样本输入到训练得到的信息推荐模型中,得到各个待预测样本与所述目标服务事件的匹配度;
 - 针对每个待预测样本,根据该待预测样本与所述目标服务事件的匹配度确定是否需要该待预测样本的服务请求方终端推荐与所述目标服务事件相关的信息。
2. 根据权利要求1所述的信息推荐方法,其特征在于,所述获得训练样本集的步骤,包括:
 - 获得多个初始正样本和多个初始未知样本;
 - 对每个初始正样本和每个初始未知样本的样本特征进行归一化处理,得到处于同一值空间范围内的多个正样本和多个未知样本。
3. 根据权利要求2所述的信息推荐方法,其特征在于,所述获得多个初始正样本和多个初始未知样本的步骤,包括:
 - 获得与所述目标服务事件匹配的多个第一用户的第一服务记录信息以及所述第一用户之外的多个第二用户的第二服务记录信息;
 - 从所述第一服务记录信息中提取特征信息,以得到多个初始正样本,并从所述第二服务记录信息中提取特征信息,以得到多个初始未知样本。
4. 根据权利要求2所述的信息推荐方法,其特征在于,所述对每个初始正样本和每个初始未知样本的样本特征进行归一化处理,得到处于同一值空间范围内的多个正样本和多个未知样本的步骤,包括:
 - 获取每个初始正样本和每个初始未知样本的各个样本特征的缺省值;
 - 对缺省值低于第一缺省值阈值的样本特征进行缺省值填充,并剔除缺省值大于第二缺省值阈值的样本特征,得到处理后的多个初始正样本和多个初始未知样本;
 - 将处理得到的多个初始正样本和多个初始未知样本映射到同一值空间范围内,得到处于同一值空间范围内的多个正样本和多个未知样本。
5. 根据权利要求1所述的信息推荐方法,其特征在于,所述基于所述多个正样本和多个未知样本,经过至少一个迭代周期的重心聚类,得到与目标服务事件不匹配的多个负样本的步骤,包括:
 - 针对每个迭代周期,经过该迭代周期内的迭代聚类,得到一个聚类后的正簇和负簇;
 - 在该正簇的多个正样本和该负簇的多个负样本中,计算每个样本与所述聚类后的正簇的重心之间的第一距离以及与聚类后的负簇的重心之间的第二距离;
 - 遍历每个样本,若样本计算得到的第一距离小于第二距离,则将该样本作为正样本,若第一距离大于第二距离,则将该样本作为负样本。
6. 根据权利要求5所述的信息推荐方法,其特征在于,所述每个迭代周期,经过该迭代周期内的迭代聚类,得到一个聚类后的正簇和负簇的步骤,包括:
 - 在第一个迭代周期中,计算所述多个正样本的重心;
 - 将与计算得到的所述重心之间的距离小于第一预设阈值的所有样本作为聚类后的第

一正簇,将与所述重心之间的距离大于第二预设阈值的所有样本作为聚类后的第一负簇。

7. 根据权利要求1-6中任意一项所述的信息推荐方法,其特征在于,基于所述多个正样本和多个未知样本,经过至少一个迭代周期的重心聚类,得到与目标服务事件不匹配的多个负样本的步骤,包括:

将与所述多个正样本的重心之间的距离小于第一预设阈值的所有样本作为聚类后的第一正簇,将与所述多个正样本的重心之间的距离大于第二预设阈值的所有样本作为聚类后的第一负簇;

在该第一正簇的多个正样本和该第一负簇的多个负样本中,计算每个样本与该第一正簇的重心之间的第一距离以及与该第一负簇的重心之间的第二距离;

遍历每个样本,若该样本计算得到的第一距离小于第二距离,则将该样本作为正样本,若第一距离大于第二距离,则将该样本作为负样本,以得到对应的第二正簇和第二负簇;

在该第二正簇的多个正样本和该第二负簇的多个负样本中,计算每个样本与该第二正簇的重心之间的第一距离以及与该第二负簇的重心之间的第二距离;

遍历每个样本,若该样本计算得到的第一距离小于第二距离,则将该样本作为正样本,若第一距离大于第二距离,则将该样本作为负样本,以得到对应的第三正簇和第三负簇;

将所述第三正簇作为新的第二正簇,将所述第三负簇作为新的第二负簇,并返回所述在该第二正簇的多个正样本和该第二负簇的多个负样本中,计算每个样本与该第二正簇的重心之间的第一距离以及与该第二负簇的重心之间的第二距离的步骤,直到满足迭代停止条件,将最后得到的第三负簇中的各个负样本作为与目标服务事件不匹配的负样本。

8. 根据权利要求7所述的信息推荐方法,其特征在于,所述迭代停止条件包括以下条件中的至少一种:

所述第三正簇中的正样本和所述第三负簇中的负样本不再发生变化;

迭代次数达到设定次数;

所述第三正簇和所述第三负簇的重心移动距离小于设定距离。

9. 根据权利要求1所述的信息推荐方法,其特征在于,所述根据所述多个正样本和所述多个负样本训练得到信息推荐模型的步骤,包括:

根据所述多个正样本和所述多个负样本训练支持向量机模型,并将训练得到的支持向量机模型作为所述信息推荐模型。

10. 根据权利要求1所述的信息推荐方法,其特征在于,所述根据所述多个正样本和所述多个负样本训练得到信息推荐模型的步骤,包括:

根据所述多个正样本和所述多个负样本训练支持向量机模型;

根据训练完成的支持向量机模型对所述多个正样本和所述多个负样本进行样本选择,得到多个目标正样本和多个目标负样本;

基于所述多个目标正样本和多个目标负样本对集成树模型进行训练,并将训练得到的集成树模型作为训练后的信息推荐模型。

11. 根据权利要求9或10所述的信息推荐方法,其特征在于,所述根据所述多个正样本和所述多个负样本训练支持向量机模型的步骤,包括:

通过预先设定的核函数对所述多个正样本和所述多个负样本进行映射变换;

将映射变换后的每个正样本和每个负样本输入到支持向量机模型中进行训练,计算映

射变换后的每个正样本和每个负样本与超平面之间的距离；

根据计算得到的每个正样本和每个负样本与超平面之间的距离计算所述支持向量机模型的损失函数值；

根据计算得到的损失函数值，利用梯度下降法对所述支持向量机模型的模型参数和学习率进行调整，并基于调整后的支持向量机模型返回将映射变换后的每个正样本和每个负样本训练输入到支持向量机模型进行训练，计算映射变换后的每个正样本和每个负样本与超平面之间的距离的步骤，直到满足迭代停止条件，输出训练得到的支持向量机模型。

12. 根据权利要求10所述的信息推荐方法，其特征在于，所述根据训练完成的支持向量机模型对所述多个正样本和所述多个负样本进行样本选择，得到多个目标正样本和多个目标负样本的步骤，包括：

将每个正样本和每个负样本输入到训练完成的支持向量机模型中，得到每个正样本和每个负样本与所述目标服务事件的匹配率；

根据每个正样本和每个负样本与所述目标服务事件的匹配率将误差值大于第三预设阈值的样本剔除，得到所述多个目标正样本和多个目标负样本。

13. 根据权利要求1所述的信息推荐方法，其特征在于，所述根据该待预测样本与所述目标服务事件的匹配度确定是否需要向该待预测样本的服务请求方终端推荐与所述目标服务事件相关的信息的步骤，包括：

判断该待预测样本与所述目标服务事件的匹配度是否大于第四预设阈值；

若该待预测样本与所述目标服务事件的匹配度大于第四预设阈值，则确定需要向该待预测样本的服务请求方终端推荐与所述目标服务事件相关的信息。

14. 根据权利要求13所述的信息推荐方法，其特征在于，所述方法还包括：

若确定需要向该待预测样本的服务请求方终端推荐与所述目标服务事件相关的信息，将该待预测样本添加到所述训练样本中的多个正样本中。

15. 根据权利要求13所述的信息推荐方法，其特征在于，所述方法还包括：

若该待预测样本与所述目标服务事件的匹配度大于第五预设阈值且不大于所述第四预设阈值，则向所述待预测样本的服务请求方终端发送提示信息以提示该服务请求方终端的用户选择是否接收所述目标服务事件相关的信息；

若接收到该服务请求方终端发送的不接收所述目标服务事件相关的信息的第一提示信息，则确定不需要向该待预测样本推荐与所述目标服务事件相关的信息，并将该待预测样本添加到与目标服务事件不匹配的多个负样本中；

若接收到该服务请求方终端发送的接收所述目标服务事件相关的信息的第二提示信息，则确定需要向该待预测样本推荐与所述目标服务事件相关的信息，并将该待预测样本添加到所述训练样本中的多个正样本中；以及

若该待预测样本与所述目标服务事件的匹配度不大于第五预设阈值，则确定不需要向该待预测样本推荐与所述目标服务事件相关的信息，并将该待预测样本添加到与目标服务事件不匹配的多个负样本中。

16. 根据权利要求1所述的信息推荐方法，其特征在于，所述根据该待预测样本与所述目标服务事件的匹配度确定是否需要向该待预测样本的服务请求方终端推荐与所述目标服务事件相关的信息的步骤之后，所述方法包括：

若确定需要向该待预测样本的服务请求方终端推荐与所述目标服务事件相关的信息，则向各个提供所述目标服务事件的服务的第三方服务器发送信息获取请求；

接收各个第三方服务器响应所述信息获取请求发送的与所述目标服务事件相关的信息，并将与所述目标服务事件相关的信息推荐给该待预测样本的服务请求方终端。

17. 一种信息推荐装置，其特征在于，应用于服务器，所述装置包括：

获得模块，用于获得训练样本集，所述训练样本集包括多个正样本和多个未知样本；

重心聚类模块，用于基于所述多个正样本和多个未知样本，经过至少一个迭代周期的重心聚类，得到与目标服务事件不匹配的多个负样本，其中，每个正样本与所述目标服务事件匹配；

训练模块，用于根据所述多个正样本和所述多个负样本训练得到信息推荐模型，并将各个待预测样本输入到训练得到的信息推荐模型中，得到各个待预测样本与所述目标服务事件的匹配度；

确定模块，用于针对每个待预测样本，根据该待预测样本与所述目标服务事件的匹配度确定是否需要向该待预测样本的服务请求方终端推荐与所述目标服务事件相关的信息。

18. 根据权利要求17所述的信息推荐装置，其特征在于，所述获得模块，具体通过以下方式获得所述多个正样本和多个未知样本：

获得多个初始正样本和多个初始未知样本；

对每个初始正样本和每个初始未知样本的样本特征进行归一化处理，得到处于同一值空间范围内的多个正样本和多个未知样本。

19. 根据权利要求18所述的信息推荐装置，其特征在于，所述获得模块，具体通过以下方式获得所述多个初始正样本和多个初始未知样本：

获得与所述目标服务事件匹配的多个第一用户的第一服务记录信息以及所述第一用户之外的多个第二用户的第二服务记录信息；

从所述第一服务记录信息中提取特征信息，以得到多个初始正样本，并从所述第二服务记录信息中提取特征信息，以得到多个初始未知样本。

20. 根据权利要求18所述的信息推荐装置，其特征在于，所述获得模块，具体通过以下方式得到处于同一值空间范围内的多个正样本和多个未知样本：

获取每个初始正样本和每个初始未知样本的各个样本特征的缺省值；

对缺省值低于第一缺省值阈值的样本特征进行缺省值填充，并剔除缺省值大于第二缺省值阈值的样本特征，得到处理后的多个初始正样本和多个初始未知样本；

将处理得到的多个初始正样本和多个初始未知样本映射到同一值空间范围内，得到处于同一值空间范围内的多个正样本和多个未知样本。

21. 根据权利要求17所述的信息推荐装置，其特征在于，所述重心聚类模块，具体用于通过以下方式得到与目标服务事件不匹配的多个负样本的：

针对每个迭代周期，经过该迭代周期内的迭代聚类，得到一个聚类后的正簇和负簇；

在该正簇的多个正样本和该负簇的多个负样本中，计算每个样本与所述聚类后的正簇的重心之间的第一距离以及与聚类后的负簇的重心之间的第二距离；

遍历每个样本，若样本计算得到的第一距离小于第二距离，则将该样本作为正样本，若第一距离大于第二距离，则将该样本作为负样本。

22. 根据权利要求21所述的信息推荐装置,其特征在于,所述重心聚类模块,具体用于通过以下方式执行第一个迭代周期的重心聚类:

在第一个迭代周期中,计算所述多个正样本的重心;

将与计算得到的所述重心之间的距离小于第一预设阈值的所有样本作为聚类后的第一正簇,将与所述重心之间的距离大于第二预设阈值的所有样本作为聚类后的第一负簇。

23. 根据权利要求17-22中任意一项所述的信息推荐装置,其特征在于,所述重心聚类模块,具体用于通过以下方式得到与目标服务事件不匹配的多个负样本的步骤,包括:

将与所述多个正样本的重心之间的距离小于第一预设阈值的所有样本作为聚类后的第一正簇,将与所述多个正样本的重心之间的距离大于第二预设阈值的所有样本作为聚类后的第一负簇;

在该第一正簇的多个正样本和该第一负簇的多个负样本中,计算每个样本与该第一正簇的重心之间的第一距离以及与该第一负簇的重心之间的第二距离;

遍历每个样本,若该样本计算得到的第一距离小于第二距离,则将该样本作为正样本,若第一距离大于第二距离,则将该样本作为负样本,以得到对应的第二正簇和第二负簇;

在该第二正簇的多个正样本和该第二负簇的多个负样本中,计算每个样本与该第二正簇的重心之间的第一距离以及与该第二负簇的重心之间的第二距离;

遍历每个样本,若该样本计算得到的第一距离小于第二距离,则将该样本作为正样本,若第一距离大于第二距离,则将该样本作为负样本,以得到对应的第三正簇和第三负簇;

将所述第三正簇作为新的第二正簇,将所述第三负簇作为新的第二负簇,并返回所述在该第二正簇的多个正样本和该第二负簇的多个负样本中,计算每个样本与该第二正簇的重心之间的第一距离以及与该第二负簇的重心之间的第二距离的步骤,直到满足迭代停止条件,将最后得到的第三负簇中的各个负样本作为与目标服务事件不匹配的负样本。

24. 根据权利要求23所述的信息推荐装置,其特征在于,所述迭代停止条件包括以下条件中的至少一种:

所述第三正簇中的正样本和所述第三负簇中的负样本不再发生变化;

迭代次数达到设定次数;

所述第三正簇和所述第三负簇的重心移动距离小于设定距离。

25. 根据权利要求17所述的信息推荐装置,其特征在于,所述训练模块,具体通过以下方式训练得到信息推荐模型:

根据所述多个正样本和所述多个负样本训练支持向量机模型,并将训练得到的支持向量机模型作为所述信息推荐模型。

26. 根据权利要求17所述的信息推荐装置,其特征在于,所述训练模块,具体通过以下方式训练得到信息推荐模型:

根据所述多个正样本和所述多个负样本训练支持向量机模型;

根据训练完成的支持向量机模型对所述多个正样本和所述多个负样本进行样本选择,得到多个目标正样本和多个目标负样本;

基于所述多个目标正样本和多个目标负样本对集成树模型进行训练,并将训练得到的集成树模型作为训练后的信息推荐模型。

27. 根据权利要求25或26所述的信息推荐装置,其特征在于,所述训练模块,具体通过

以下方式训练支持向量机模型：

通过预先设定的核函数对所述多个正样本和所述多个负样本进行映射变换；

将映射变换后的每个正样本和每个负样本输入到支持向量机模型中进行训练，计算映射变换后的每个正样本和每个负样本与超平面之间的距离；

根据计算得到的每个正样本和每个负样本与超平面之间的距离计算所述支持向量机模型的损失函数值；

根据计算得到的损失函数值，利用梯度下降法对所述支持向量机模型的模型参数和学习率进行调整，并基于调整后的支持向量机模型返回将映射变换后的每个正样本和每个负样本训练输入到支持向量机模型进行训练，计算映射变换后的每个正样本和每个负样本与超平面之间的距离的步骤，直到满足迭代停止条件，输出训练得到的支持向量机模型。

28. 根据权利要求27所述的信息推荐装置，其特征在于，所述训练模块，具体通过以下方式得到多个目标正样本和多个目标负样本：

将每个正样本和每个负样本输入到训练完成的支持向量机模型中，得到每个正样本和每个负样本与所述目标服务事件的匹配率；

根据每个正样本和每个负样本与所述目标服务事件的匹配率将误差值大于第三预设阈值的样本剔除，得到所述多个目标正样本和多个目标负样本。

29. 根据权利要求17所述的信息推荐装置，其特征在于，所述确定模块通过以下方式确定是否需要向该待预测样本的服务请求方终端推荐与所述目标服务事件相关的信息：

判断该待预测样本与所述目标服务事件的匹配度是否大于第四预设阈值；

若该待预测样本与所述目标服务事件的匹配度大于第四预设阈值，则确定需要向该待预测样本的服务请求方终端推荐与所述目标服务事件相关的信息。

30. 根据权利要求17所述的信息推荐装置，其特征在于，若确定需要向该待预测样本的服务请求方终端推荐与所述目标服务事件相关的信息，所述确定模块，具体还用于：

将该待预测样本添加到所述训练样本中的多个正样本中。

31. 根据权利要求29所述的信息推荐装置，其特征在于，若该待预测样本与所述目标服务事件的匹配度不大于所述第四预设阈值，则所述确定模块，具体还用于：

向所述待预测样本的服务请求方终端发送提示信息以提示该服务请求方终端的用户选择是否接收所述目标服务事件相关的信息；

若接收到该服务请求方终端发送的不接收所述目标服务事件相关的信息的第一指示信息，则确定不需要向该待预测样本推荐与所述目标服务事件相关的信息，并将该待预测样本添加到与目标服务事件不匹配的多个负样本中；

若接收到该服务请求方终端发送的接收所述目标服务事件相关的信息的第二指示信息，则确定需要向该待预测样本推荐与所述目标服务事件相关的信息，并将该待预测样本添加到所述训练样本中的多个正样本中。

32. 根据权利要求17所述的信息推荐装置，其特征在于，若确定需要向该待预测样本的服务请求方终端推荐与所述目标服务事件相关的信息，所述装置还包括：

信息推荐模块，用于向各个提供所述目标服务事件的服务的第三方服务器发送信息获取请求，接收各个第三方服务器响应所述信息获取请求发送的与所述目标服务事件相关的信息，并将与所述目标服务事件相关的信息推荐给该待预测样本的服务请求方终端。

33. 一种服务器,其特征在于,包括:处理器、存储介质和总线,所述存储介质存储有所述处理器可执行的机器可读指令,当服务器运行时,所述处理器与所述存储介质之间通过总线通信,所述处理器执行所述机器可读指令,以执行时执行如权利要求1-16中任意一项所述的信息推荐方法的步骤。

34. 一种可读存储介质,其特征在于,该可读存储介质上存储有计算机程序,该计算机程序被处理器运行时执行如权利要求1-16中任意一项所述的信息推荐方法的步骤。

信息推荐方法、装置、服务器及可读存储介质

技术领域

[0001] 本申请涉及计算机技术领域,具体而言,涉及一种信息推荐方法、装置、服务器及可读存储介质。

背景技术

[0002] 目前,随着智能终端的普及,各种提供生活便利服务的应用程序(Application,简称APP)也层出不穷,为用户的吃穿住行提供服务。这些应用程序每天都能收集到大量用户的服务记录信息,这些服务记录信息包含了用户的习惯和喜好,也可以用来预测用户与该应用程序提供的服务无关的行为。然而,如何从大量用户的服务记录信息中确定每个用户是否对各类预推荐的目标服务感兴趣,从而有针对性地为这些用户推荐匹配的目标服务相关的信息,是本领域技术人员亟待解决的技术问题。

发明内容

[0003] 有鉴于此,本申请实施例的目的在于提供一种信息推荐方法、装置、服务器及可读存储介质,能够有针对性地为服务请求方推荐匹配的目标服务相关的信息。

[0004] 根据本申请实施例的一个方面,提供一种电子设备,可以包括一个或多个存储介质和一个或多个与存储介质通信的处理器。一个或多个存储介质存储有处理器可执行的机器可读指令。当电子设备运行时,处理器与存储介质之间通过总线通信,处理器执行所述机器可读指令,以执行信息推荐方法。

[0005] 根据本申请实施例的另一方面,提供一种信息推荐方法,应用于服务器,所述方法可以包括:

[0006] 获得训练样本集,所述训练样本集包括多个正样本和多个未知样本;

[0007] 基于所述多个正样本和多个未知样本,经过至少一个迭代周期的重心聚类,得到与目标服务事件不匹配的多个负样本,其中,每个正样本与所述目标服务事件匹配;

[0008] 根据所述多个正样本和所述多个负样本训练得到信息推荐模型,并将各个待预测样本输入到训练得到的信息推荐模型中,得到各个待预测样本与所述目标服务事件的匹配度;

[0009] 针对每个待预测样本,根据该待预测样本与所述目标服务事件的匹配度确定是否需要向该待预测样本的服务请求方终端推荐与所述目标服务事件相关的信息。

[0010] 在一种可能的实施方式中,所述获得训练样本集的步骤,可以包括:

[0011] 获得多个初始正样本和多个初始未知样本;

[0012] 对每个初始正样本和每个初始未知样本的样本特征进行归一化处理,得到处于同一值空间范围内的多个正样本和多个未知样本。

[0013] 在一种可能的实施方式中,所述获得多个初始正样本和多个初始未知样本的步骤,可以包括:

[0014] 获得与所述目标服务事件匹配的多个第一用户的第一服务记录信息以及所述第

一用户之外的多个第二用户的第二服务记录信息；

[0015] 从所述第一服务记录信息中提取特征信息,以得到多个初始正样本,并从所述第二服务记录信息中提取特征信息,以得到多个初始未知样本。

[0016] 在一种可能的实施方式中,所述对每个初始正样本和每个初始未知样本的样本特征进行归一化处理,得到处于同一值空间范围内的多个正样本和多个未知样本的步骤,可以包括:

[0017] 获取每个初始正样本和每个初始未知样本的各个样本特征的缺省值;

[0018] 对缺省值低于第一缺省值阈值的样本特征进行缺省值填充,并剔除缺省值大于第二缺省值阈值的样本特征,得到处理后的多个初始正样本和多个初始未知样本;

[0019] 将处理得到的多个初始正样本和多个初始未知样本映射到同一值空间范围内,得到处于同一值空间范围内的多个正样本和多个未知样本。

[0020] 在一种可能的实施方式中,所述基于所述多个正样本和多个未知样本,经过至少一个迭代周期的重心聚类,得到与目标服务事件不匹配的多个负样本的步骤,可以包括:

[0021] 针对每个迭代周期,经过该迭代周期内的迭代聚类,得到一个聚类后的正簇和负簇;

[0022] 在该正簇的多个正样本和该负簇的多个负样本中,计算每个样本与所述聚类后的正簇的重心之间的第一距离以及与聚类后的负簇的重心之间的第二距离;

[0023] 遍历每个样本,若样本计算得到的第一距离小于第二距离,则将该样本作为正样本,若第一距离大于第二距离,则将该样本作为负样本。

[0024] 在一种可能的实施方式中,所述每个迭代周期,经过该迭代周期内的迭代聚类,得到一个聚类后的正簇和负簇的步骤,可以包括:

[0025] 在第一个迭代周期中,计算所述多个正样本的重心;

[0026] 将与计算得到的所述重心之间的距离小于第一预设阈值的所有样本作为聚类后的第一正簇,将与所述重心之间的距离大于第二预设阈值的所有样本作为聚类后的第一负簇。

[0027] 在一种可能的实施方式中,基于所述多个正样本和多个未知样本,经过至少一个迭代周期的重心聚类,得到与目标服务事件不匹配的多个负样本的步骤,可以包括:

[0028] 将与所述多个正样本的重心之间的距离小于第一预设阈值的所有样本作为聚类后的第一正簇,将与所述多个正样本的重心之间的距离大于第二预设阈值的所有样本作为聚类后的第一负簇;

[0029] 在该第一正簇的多个正样本和该第一负簇的多个负样本中,计算每个样本与该第一正簇的重心之间的第一距离以及与该第一负簇的重心之间的第二距离;

[0030] 遍历每个样本,若该样本计算得到的第一距离小于第二距离,则将该样本作为正样本,若第一距离大于第二距离,则将该样本作为负样本,以得到对应的第二正簇和第二负簇;

[0031] 在该第二正簇的多个正样本和该第二负簇的多个负样本中,计算每个样本与该第二正簇的重心之间的第一距离以及与该第二负簇的重心之间的第二距离;

[0032] 遍历每个样本,若该样本计算得到的第一距离小于第二距离,则将该样本作为正样本,若第一距离大于第二距离,则将该样本作为负样本,以得到对应的第三正簇和第三负

簇；

[0033] 将所述第三正簇作为新的第二正簇,将所述第三负簇作为新的第二负簇,并返回所述在该第二正簇的多个正样本和该第二负簇的多个负样本中,计算每个样本与该第二簇的重心之间的第一距离以及与该第二负簇的重心之间的第二距离的步骤,直到满足迭代停止条件,将最后得到的第三负簇中的各个负样本作为与目标服务事件不匹配的负样本。

[0034] 在一种可能的实施方式中,所述迭代停止条件包括以下条件中的至少一种:

[0035] 所述第三正簇中的正样本和所述第三负簇中的负样本不再发生变化;

[0036] 迭代次数达到设定次数;

[0037] 所述第三正簇和所述第三负簇的重心移动距离小于设定距离。

[0038] 在一种可能的实施方式中,所述根据所述多个正样本和所述多个负样本训练得到信息推荐模型的步骤,可以包括:

[0039] 根据所述多个正样本和所述多个负样本训练支持向量机模型,并将训练得到的支持向量机模型作为所述信息推荐模型。

[0040] 在一种可能的实施方式中,所述根据所述多个正样本和所述多个负样本训练得到信息推荐模型的步骤,可以包括:

[0041] 根据所述多个正样本和所述多个负样本训练支持向量机模型;

[0042] 根据训练完成的支持向量机模型对所述多个正样本和所述多个负样本进行样本选择,得到多个目标正样本和多个目标负样本;

[0043] 基于所述多个目标正样本和多个目标负样本对集成树模型进行训练,并将训练得到的集成树模型作为训练后的信息推荐模型。

[0044] 在一种可能的实施方式中,所述根据所述多个正样本和所述多个负样本训练支持向量机模型的步骤,可以包括:

[0045] 通过预先设定的核函数对所述多个正样本和所述多个负样本进行映射变换;

[0046] 将映射变换后的每个正样本和每个负样本输入到支持向量机模型中进行训练,计算映射变换后的每个正样本和每个负样本与超平面之间的距离;

[0047] 根据计算得到的每个正样本和每个负样本与超平面之间的距离计算所述支持向量机模型的损失函数值;

[0048] 根据计算得到的损失函数值,利用梯度下降法对所述支持向量机模型的模型参数和学习率进行调整,并基于调整后的支持向量机模型返回将映射变换后的每个正样本和每个负样本训练输入到支持向量机模型进行训练,计算映射变换后的每个正样本和每个负样本与超平面之间的距离的步骤,直到满足迭代停止条件,输出训练得到的支持向量机模型。

[0049] 在一种可能的实施方式中,所述根据训练完成的支持向量机模型对所述多个正样本和所述多个负样本进行样本选择,得到多个目标正样本和多个目标负样本的步骤,可以包括:

[0050] 将每个正样本和每个负样本输入到训练完成的支持向量机模型中,得到每个正样本和每个负样本与所述目标服务事件的匹配率;

[0051] 根据每个正样本和每个负样本与所述目标服务事件的匹配率将误差值大于第三预设阈值的样本剔除,得到所述多个目标正样本和多个目标负样本。

[0052] 在一种可能的实施方式中,所述根据该待预测样本与所述目标服务事件的匹配度

确定是否需要向该待预测样本的服务请求方终端推荐与所述目标服务事件相关的信息的步骤,可以包括:

[0053] 判断该待预测样本与所述目标服务事件的匹配度是否大于第四预设阈值;

[0054] 若该待预测样本与所述目标服务事件的匹配度大于第四预设阈值,则确定需要向该待预测样本的服务请求方终端推荐与所述目标服务事件相关的信息。

[0055] 在一种可能的实施方式中,所述方法还可以包括:

[0056] 若确定需要向该待预测样本的服务请求方终端推荐与所述目标服务事件相关的信息,将该待预测样本添加到所述训练样本中的多个正样本中。

[0057] 在一种可能的实施方式中,所述方法还可以包括:

[0058] 若该待预测样本与所述目标服务事件的匹配度大于第五预设阈值且不大于所述第四预设阈值,则向所述待预测样本的服务请求方终端发送提示信息以提示该服务请求方终端的用户选择是否接收所述目标服务事件相关的信息;

[0059] 若接收到该服务请求方终端发送的不接收所述目标服务事件相关的信息的第一指示信息,则确定不需要向该待预测样本推荐与所述目标服务事件相关的信息,并将该待预测样本添加到与目标服务事件不匹配的多个负样本中;

[0060] 若接收到该服务请求方终端发送的接收所述目标服务事件相关的信息的第二指示信息,则确定需要向该待预测样本推荐与所述目标服务事件相关的信息,并将该待预测样本添加到所述训练样本中的多个正样本中;以及

[0061] 若该待预测样本与所述目标服务事件的匹配度不大于第五预设阈值,则确定不需要向该待预测样本推荐与所述目标服务事件相关的信息,并将该待预测样本添加到与目标服务事件不匹配的多个负样本中。

[0062] 在一种可能的实施方式中,所述根据该待预测样本与所述目标服务事件的匹配度确定是否需要向该待预测样本的服务请求方终端推荐与所述目标服务事件相关的信息的步骤之后,所述方法包括:

[0063] 若确定需要向该待预测样本的服务请求方终端推荐与所述目标服务事件相关的信息,则向各个提供所述目标服务事件的服务的第三方服务器发送信息获取请求;

[0064] 接收各个第三方服务器响应所述信息获取请求发送的与所述目标服务事件相关的信息,并将与所述目标服务事件相关的信息推荐给该待预测样本的服务请求方终端。

[0065] 根据本申请实施例的另一方面,提供一种信息推荐装置,应用于服务器,所述装置可以包括:

[0066] 获得模块,用于获得训练样本集,所述训练样本集包括多个正样本和多个未知样本;

[0067] 重心聚类模块,用于基于所述多个正样本和多个未知样本,经过至少一个迭代周期的重心聚类,得到与目标服务事件不匹配的多个负样本,其中,每个正样本与所述目标服务事件匹配;

[0068] 训练模块,用于根据所述多个正样本和所述多个负样本训练得到信息推荐模型,并将各个待预测样本输入到训练得到的信息推荐模型中,得到各个待预测样本与所述目标服务事件的匹配度;

[0069] 确定模块,用于针对每个待预测样本,根据该待预测样本与所述目标服务事件的

匹配度确定是否需要向该待预测样本的服务请求方终端推荐与所述目标服务事件相关的信息。

[0070] 根据本申请实施例的另一方面,提供一种可读存储介质,该可读存储介质上存储有计算机程序,该计算机程序被处理器运行时可以执行上述的信息推荐方法的步骤。

[0071] 基于上述任一方面,本申请实施例通过对训练样本集中的多个正样本和多个未知样本经过至少一个迭代周期的重心聚类,可以得到更为准确的多个与目标服务事件不匹配的负样本,避免在训练信息推荐模型过程中错误地将原本与目标服务事件匹配的样本作为负样本的特征信息进行学习而导致的预测误差。然后按照以此训练得到的信息推荐模型预测各个待预测样本与目标服务事件的匹配度,从而确定是否需要向各个待预测样本的服务请求方终端推荐与目标服务事件相关的信息,如此,能够提高信息推荐的精准性,同时为服务请求方提供更多便利、高效的目标服务。

[0072] 为使本申请实施例的上述目的、特征和优点能更明显易懂,下面将结合实施例,并配合所附附图,作详细说明。

附图说明

[0073] 为了更清楚地说明本申请实施例的技术方案,下面将对实施例中所需要使用的附图作简单地介绍,应当理解,以下附图仅示出了本申请的某些实施例,因此不应被看作是对范围的限定,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他相关的附图。

[0074] 图1示出了本申请实施例所提供的信息推荐系统的交互示意框图;

[0075] 图2示出了本申请实施例所提供的可以实现图1的服务器、服务请求方终端、服务提供方终端的电子设备的示例性硬件和软件组件的示意图;

[0076] 图3示出了本申请实施例所提供的信息推荐方法的流程示意图之一;

[0077] 图4为图3示出的信息推荐方法中步骤S120的流程示意图;

[0078] 图5示出了本申请实施例所提供的信息推荐方法的流程示意图之二;

[0079] 图6示出了本申请实施例所提供的信息推荐装置的功能模块框图之一;

[0080] 图7示出了本申请实施例所提供的信息推荐装置的功能模块框图之二。

具体实施方式

[0081] 为使本申请实施例的目的、技术方案和优点更加清楚,下面将结合本申请实施例中的附图,对本申请实施例中的技术方案进行清楚、完整地描述,应当理解,本申请中附图仅起到说明和描述的目的,并不用于限定本申请的保护范围。另外,应当理解,示意性的附图并未按实物比例绘制。本申请中使用的流程图示出了根据本申请实施例的一些实施例实现的操作。应该理解,流程图的操作可以不按顺序实现,没有逻辑的上下文关系的步骤可以反转顺序或者同时实施。此外,本领域技术人员在本申请内容的指引下,可以向流程图添加一个或多个其他操作,也可以从流程图中移除一个或多个操作。

[0082] 另外,所描述的实施例仅仅是本申请一部分实施例,而不是全部的实施例。通常在此处附图中描述和示出的本申请实施例的组件可以以各种不同的配置来布置和设计。因此,以下对在附图中提供的本申请的实施例的详细描述并非旨在限制要求保护的本申请的

范围,而是仅仅表示本申请的选定实施例。基于本申请的实施例,本领域技术人员在没有做出创造性劳动的前提下所获得的所有其他实施例,都属于本申请保护的范畴。

[0083] 为了使得本领域技术人员能够使用本申请内容,结合特定应用场景“网约车打车场景”,给出以下实施方式。对于本领域技术人员来说,在不脱离本申请的精神和范围的情况下,可以将这里定义的一般原理应用于其他实施例和应用场景。虽然本申请主要围绕“网约车打车场景”进行描述,但是应该理解,这仅是一个示例性实施例。本申请可以应用于任何其他交通运输类型。例如,本申请可以应用于不同的运输系统环境,包括陆地,海洋,或航空等,或其任意组合。运输系统的交通工具可以包括出租车、私家车、顺风车、公共汽车、火车、子弹头列车、高速铁路、地铁、船只、飞机、宇宙飞船、热气球、或无人驾驶车辆等,或其任意组合。本申请还可以包括用于网约车打车的任何服务系统,例如,用于发送和/或接收快递的系统、用于买卖双方交易的服务系统。本申请的系统或方法的应用可以包括网页、浏览器的插件、客户端终端、定制系统、内部分析系统、或人工智能机器人等,或其任意组合。

[0084] 需要说明的是,本申请实施例中将会用到术语“包括”,用于指出其后所声明的特征的存在,但并不排除增加其他的特征。

[0085] 本申请中的术语“乘客”、“请求方”、“服务人员”、“服务请求方”和“客户”可互换使用,以指代可以请求或订购服务的个人、实体或工具。本申请中的术语“司机”、“提供方”、“服务提供方”和“供应商”可互换使用,以指代可以提供服务的个人、实体或工具。本申请中的术语“用户”可以指代请求服务、订购服务、提供服务或促成服务的提供的个人、实体或工具。例如,用户可以是乘客、驾驶员、操作员等,或其任意组合。在本申请中,“乘客”和“乘客终端”可以互换使用,“驾驶员”和“驾驶员终端”可以互换使用。

[0086] 为了解决本申请背景技术中所述的至少一种技术问题,本申请发明人经过仔细研究发现,目前的做法是通过收集大量与目标服务事件匹配的正样本,然后将其余不确定是否与该目标服务事件匹配的样本都作为负样本,并根据这些收集到的正样本和负样本训练得到信息推荐模型,再通过训练得到的信息推荐模型识别每个服务请求方是否对该目标服务感兴趣,从而有针对性地为这些对该目标服务感兴趣的服务请求方推荐该目标服务相关的信息。

[0087] 在一个示例中,以网约车应用程序(例如滴滴出行)为例,网约车应用程序每天都能收集到大量的出行服务记录信息,这些出行服务记录信息包含了服务请求方(例如,乘客,司机等)的行为信息,例如使用习惯、行为偏好等等。同时,这些出行服务记录信息也可以用来预测服务请求方与网约车无关的行为,例如,出行服务记录信息可能蕴含了服务请求方有买车或者租车意图,如果能够判断这个意图,识别有买车或者租车意图的服务请求方,就可以有针对性地进行相应的信息推荐。

[0088] 针对上述的示例,目前的方案则是:首先收集大量已经有买车或者租车意图的服务请求方的出行服务记录信息,并提取这些出行服务记录信息的特征信息,从而得到正样本。同时,收集其余不确定是否有买车或者租车意图的服务请求方的出行服务记录信息,并提取这些出行服务记录信息的特征信息,从而得到负样本。接着,根据这些收集到的正样本和负样本训练得到信息推荐模型,再通过训练得到的信息推荐模型识别每个服务请求方是否有买车或者租车意图,从而有针对性地为这些对该目标服务感兴趣的服务请求方推荐买车或者租车相关的信息。

[0089] 然而,发明人在仔细研究后发现,如果认定除去正样本以外的样本都是负样本,并以此来训练得到的信息推荐模型,其预测结果会存在极大误差。例如,目前方案中所认定的负样本并不都是与目标服务事件不匹配的样本,也会包括与目标服务事件匹配的样本,由此会导致在训练信息推荐模型过程中会错误地将原本与目标服务事件匹配的样本,也即部分正样本的特征信息作为负样本的特征信息进行学习,从而必然会导致后续训练得到的信息推荐模型的预测误差。

[0090] 需要注意的是,以上现有技术中的方案所存在的缺陷,均是发明人在经过实践并仔细研究后得出的结果,因此,上述问题的发现过程以及下文中本申请实施例针对上述问题所提出的解决方案,都应该是发明人在本申请过程中对本申请做出的贡献。

[0091] 根据本申请发明人对上述技术问题的研究,本申请实施例提供一种信息推荐方法、装置、服务器及可读存储介质,通过对训练样本集中的多个正样本和多个未知样本经过至少一个迭代周期的重心聚类,可以得到更为准确的多个与目标服务事件不匹配的负样本,避免在训练信息推荐模型过程中错误地将原本与目标服务事件匹配的样本作为负样本的特征信息进行学习而导致的预测误差。然后按照以此训练得到的信息推荐模型预测各个待预测样本与目标服务事件的匹配度,从而确定是否需要向各个待预测样本的服务请求方终端推荐与目标服务事件相关的信息,如此,能够提高信息推荐的精准性,同时为服务请求方提供更多便利、高效的目标服务。

[0092] 图1是本申请一种可替代实施例提供的信息推荐系统100的架构示意图。例如,信息推荐系统100可以是用于诸如出租车、代驾服务、快车服务、拼车服务、公共汽车服务、驾驶员租赁服务、或班车服务之类的运输服务、或其上述任意服务时间的组合服务所依赖的在线运输服务平台。信息推荐系统100可以包括服务器110、网络120、服务请求方终端130、服务提供方终端140和数据库150,服务器110中可以包括执行指令操作的处理器。图1所示的信息推荐系统100仅为一种可行的示例,在其它可行的实施例中,该信息推荐系统100也可以仅包括图1所示组成部分的其中一部分或者还可以包括其它的组成部分。

[0093] 在一些实施例中,服务器110可以是单个服务器,也可以是一个服务器组。服务器组可以是集中式的,也可以是分布式的(例如,服务器110可以是分布式系统)。在一些实施例中,服务器110相对于终端,可以是本地的、也可以是远程的。例如,服务器110可以经由网络120访问存储在服务请求方终端130、服务提供方终端140以及数据库150、或其任意组合中的信息。作为另一示例,服务器110可以直接连接到服务请求方终端130、服务提供方终端140和数据库150中的至少一个,以访问其中存储的信息和/或数据。在一些实施例中,服务器110可以在云平台上实现;仅作为示例,云平台可以包括私有云、公有云、混合云、社区云(community cloud)、分布式云、跨云(inter-cloud)、多云(multi-cloud)等,或者它们的任意组合。在一些实施例中,服务器110可以在具有本申请中图2所示的一个或多个组件的电子设备200上实现。

[0094] 在一些实施例中,服务器110可以包括处理器。处理器可以处理与服务请求有关的信息和/或数据,以执行本申请中描述的一个或多个功能。例如,在快车服务中,处理器可以基于从服务请求方终端130获得的服务请求来确定目标车辆。处理器可以包括一个或多个处理核(例如,单核处理器(S)或多核处理器(S))。仅作为举例,处理器可以包括中央处理单元(Central Processing Unit,CPU)、专用集成电路(Application Specific Integrated

Circuit,ASIC)、专用指令集处理器(Application Specific Instruction-set Processor,ASIP)、图形处理单元(Graphics Processing Unit,GPU)、物理处理单元(Physics Processing Unit,PPU)、数字信号处理器(Digital Signal Processor,DSP)、现场可编程门阵列(Field Programmable Gate Array,FPGA)、可编程逻辑器件(Programmable Logic Device,PLD)、控制器、微控制器单元、简化指令集计算机(Reduced Instruction Set Computing,RISC)、或微处理器等,或其任意组合。

[0095] 网络120可以用于信息和/或数据的交换。在一些实施例中,信息推荐系统100中的一个或多个组件(例如,服务器110,服务请求方终端130,服务提供方终端140和数据库150)可以向其他组件发送信息和/或数据。例如,服务器110可以经由网络120从服务请求方终端130获取服务请求。在一些实施例中,网络120可以是任何类型的有线或者无线网络,或者是他们的结合。仅作为示例,网络130可以包括有线网络、无线网络、光纤网络、远程通信网络、内联网、因特网、局域网(Local Area Network,LAN)、广域网(Wide Area Network,WAN)、无线局域网(Wireless Local Area Networks,WLAN)、城域网(Metropolitan Area Network,MAN)、广域网(Wide Area Network,WAN)、公共电话交换网(Public Switched Telephone Network,PSTN)、蓝牙网络、ZigBee网络、或近场通信(Near Field Communication,NFC)网络等,或其任意组合。在一些实施例中,网络120可以包括一个或多个网络接入点。例如,网络120可以包括有线或无线网络接入点,例如基站和/或网络交换节点,信息推荐系统100的一个或多个组件可以通过该接入点连接到网络120以交换数据和/或信息。

[0096] 在一些实施例中,服务请求方终端130的用户可以是除服务实际需求者之外的其他人。例如,服务请求方终端130的用户A可以使用服务请求方终端130来为服务实际需求者B发起服务请求(比如,用户A可以为自己的朋友B叫车),或者从服务器110接收服务信息或指令等。在一些实施例中,服务提供方终端140的用户可以是服务实际提供者,也可以是除服务实际提供者之外的其他人。例如,服务提供方终端140的用户C可以使用服务提供方终端140接收由服务实际提供者D提供服务的服务请求(比如用户C可以为自己雇用的司机D接单),和/或来自服务器110的信息或指令。在一些实施例中,“服务请求方”和“服务请求方终端”可以互换使用,“服务提供方”和“服务提供方终端”可以互换使用。

[0097] 在一些实施例中,服务请求方终端130可以包括移动设备、平板计算机、膝上型计算机、或机动车辆中的内置设备等,或其任意组合。在一些实施例中,移动设备可以包括智能家居设备、可穿戴设备、智能移动设备、虚拟现实设备、或增强现实设备等,或其任意组合。在一些实施例中,智能家居设备可以包括智能照明设备、智能电器设备的控制设备、智能监控设备、智能电视、智能摄像机、或对讲机等,或其任意组合。在一些实施例中,可穿戴设备可包括智能手环、智能鞋带、智能玻璃、智能头盔、智能手表、智能服装、智能背包、智能配件等、或其任何组合。在一些实施例中,智能移动设备可以包括智能手机、个人数字助理(Personal Digital Assistant,PDA)、游戏设备、导航设备、或销售点(point of sale,POS)设备等,或其任意组合。在一些实施例中,虚拟现实设备和/或增强现实设备可以包括虚拟现实头盔、虚拟现实玻璃、虚拟现实贴片、增强现实头盔、增强现实玻璃、或增强现实贴片等,或其任意组合。例如,虚拟现实设备和/或增强现实设备可以包括各种虚拟现实产品等。在一些实施例中,机动车辆中的内置设备可以包括车载计算机、车载电视等。

[0098] 数据库150可以存储数据和/或指令。在一些实施例中,数据库150可以存储从服务

请求方终端130和/或服务提供方终端140获得的数据。在一些实施例中,数据库150可以存储在本申请中描述的示例性方法的数据和/或指令。在一些实施例中,数据库150可以包括大容量存储器、可移动存储器、易失性读写存储器、或只读存储器(Read-Only Memory,ROM)等,或其任意组合。作为举例,大容量存储器可以包括磁盘、光盘、固态驱动器等;可移动存储器可包括闪存驱动器、软盘、光盘、存储卡、zip磁盘、磁带等;易失性读写存储器可以包括随机存取存储器(Random Access Memory,RAM);RAM可以包括动态RAM(Dynamic Random Access Memory,DRAM),双倍数据速率同步动态RAM(Double Date-Rate Synchronous RAM,DDR SDRAM);静态RAM(Static Random-Access Memory,SRAM),晶闸管RAM(Thyristor-Based Random Access Memory,T-RAM)和零电容器RAM(Zero-RAM)等。作为举例,ROM可以包括掩模ROM(Mask Read-Only Memory,MROM)、可编程ROM(Programmable Read-Only Memory,PROM)、可擦除可编程ROM(Programmable Erasable Read-only Memory,PEROM)、电可擦除可编程ROM(Electrically Erasable Programmable read only memory,EEPROM)、光盘ROM(CD-ROM)、以及数字通用磁盘ROM等。在一些实施例中,数据库150可以在云平台上实现。仅作为示例,云平台可以包括私有云、公有云、混合云、社区云、分布式云、跨云、多云或者其它类似的等,或其任意组合。

[0099] 在一些实施例中,数据库150可以连接到网络120以与信息推荐系统100(例如,服务器110,服务请求方终端130,服务提供方终端140等)中的一个或多个组件通信。信息推荐系统100中的一个或多个组件可以经由网络120访问存储在数据库150中的数据或指令。在一些实施例中,数据库150可以直接连接到信息推荐系统100中的一个或多个组件(例如,服务器110,服务请求方终端130,服务提供方终端140等);或者,在一些实施例中,数据库150也可以是服务器110的一部分。

[0100] 在一些实施例中,信息推荐系统100中的一个或多个组件(例如,服务器110,服务请求方终端130,服务提供方终端140等)可以具有访问数据库150的权限。在一些实施例中,当满足一定条件时,信息推荐系统100中的一个或多个组件可以读取和/或修改与服务请求方、服务提供方、或公众、或其任意组合有关的信息。例如,服务器110可以在接收服务请求之后读取和/或修改一个或多个用户的信息。

[0101] 在一些实施例中,可以通过请求服务来实现信息推荐系统100中的一个或多个组件的信息交换。服务请求的对象可以是任何产品。在一些实施方案中,产品可以是有形产品或非物质产品。有形产品可包括食品、药品、商品、化学产品、电器、服装、汽车、房屋、或奢侈品等,或其任意组合。非物质产品可以包括服务产品、金融产品、知识产品、或互联网产品等,或其任意组合。互联网产品可以包括单独的主机产品、网络产品、移动互联网产品、商业主机产品、或嵌入式产品等,或其任意组合。互联网产品可以用在移动终端的软件、程序、或系统等,或者它们的任意组合中。移动终端可以包括平板电脑、笔记本电脑、移动电话、个人数字助理(Personal Digital Assistant,PDA)、智能手表、销售点(Point of sales,POS)设备、车载电脑、车载电视、或可穿戴设备等,或其任意组合。例如,互联网产品可以是计算机或移动电话中使用的任何软件和/或应用程序。软件和/或应用程序可以涉及社交、购物、运输、娱乐时间、学习、或投资等,或其任意组合。在一些实施例中,与运输有关的软件和/或应用程序可以包括旅行软件和/或应用程序、车辆调度软件和/或应用程序、绘图软件和/或应用程序等。在车辆调度软件和/或应用程序中,车辆可包括马、马车、人力车(例如,独轮

车、自行车、三轮车等)、汽车(例如,出租车、公共汽车、私家车等)、火车、地铁、船只、飞机(例如,飞机、直升机、航天飞机、火箭、热气球等)等,或其任意组合。

[0102] 图2示出了本申请一些实施例提供的可以实现本申请思想的服务器110、服务请求方终端130、服务提供方终端140的电子设备200的示例性硬件和软件组件的示意图。例如,处理器220可以用于电子设备200上,并且用于执行本申请中的功能。

[0103] 电子设备200可以是通用计算机或特殊用途的计算机,两者都可以用于实现本申请的信息推荐方法。本申请尽管仅示出了一个计算机,但是为了方便起见,可以在多个类似平台上以分布式方式实现本申请描述的功能,以均衡处理负载。

[0104] 例如,电子设备200可以包括连接到网络的网络端口210、用于执行程序指令的一个或多个处理器220、通信总线230、和不同形式的存储介质240,例如,磁盘、ROM、或RAM,或其任意组合。示例性地,计算机平台还可以包括存储在ROM、RAM、或其他类型的非暂时性存储介质、或其任意组合中的程序指令。根据这些程序指令可以实现本申请的方法。电子设备200还包括计算机与其他输入输出设备(例如键盘、显示屏)之间的输入/输出(Input/Output, I/O)接口250。

[0105] 为了便于说明,在电子设备200中仅描述了一个处理器。然而,应当注意,本申请中的电子设备200还可以包括多个处理器,因此本申请中描述的一个处理器执行的步骤也可以由多个处理器联合执行或单独执行。例如,若电子设备200的处理器执行步骤A和步骤B,则应该理解,步骤A和步骤B也可以由两个不同的处理器共同执行或者在一个处理器中单独执行。例如,第一处理器执行步骤A,第二处理器执行步骤B,或者第一处理器和第二处理器共同执行步骤A和B。

[0106] 图3示出了本申请一些实施例提供的信息推荐方法的流程示意图,该信息推荐方法可由图1中所示的服务器110执行。应当理解,在其它实施例中,本实施例所述的信息推荐方法其中部分步骤的顺序可以根据实际需要相互交换,或者其中的部分步骤也可以省略或删除。该信息推荐方法的详细步骤介绍如下。

[0107] 步骤S110,获得训练样本集。

[0108] 本实施例中,训练样本集可包括多个正样本和多个未知样本。在一种可替换的实施方式中,本步骤S110可以通过如下方式获得训练样本集:

[0109] 首先,获得多个初始正样本和多个初始未知样本。例如,可以获得与目标服务事件匹配的多个第一用户的第一服务记录信息以及第一用户之外的多个第二用户的第二服务记录信息,然后从第一服务记录信息中提取特征信息,以得到多个初始正样本,并从第二服务记录信息中提取特征信息,以得到多个初始未知样本。

[0110] 以网约车应用程序为例,假设该目标服务事件为买车或者租车,那么所有有买车或者租车意图的用户均可以作为多个第一用户,在此基础上确定有买车或者租车意图的第一用户以及第一用户之外的多个第二用户后,从每个第一用户生成的第一出行订单(例如打快车、打优享车等生成的出行订单)中提取每个第一用户的特征信息(例如使用习惯、行为偏好、年龄、驾龄、性别等等)作为多个初始正样本。同时,从每个第二用户生成的第二出行订单中提取每个第二用户的特征信息,作为多个初始未知样本。

[0111] 其中,上述的第一用户或者第二用户可以是使用该应用程序的任何用户,仍旧以网约车应用程序为例,上述的第一用户或者第二用户可以是司机、乘客、单车用户等。

[0112] 在上述基础上,对每个初始正样本和每个初始未知样本的样本特征进行归一化处理,得到处于同一值空间范围内的多个正样本和多个未知样本。

[0113] 作为一种实施方式,可以在对每个初始正样本和每个初始未知样本的样本特征进行归一化处理之前,获取每个初始正样本和每个初始未知样本的各个样本特征的缺省值。接着,对缺省值低于第一缺省值阈值的样本特征进行缺省值填充,并剔除缺省值大于第二缺省值阈值的样本特征,得到处理后的多个初始正样本和多个初始未知样本。如此设计,能够剔除初始正样本和初始未知样本中存在的噪声样本,避免后续这些噪声样本造成训练误差。

[0114] 其中,对缺省值低于第一缺省值阈值的样本特征进行缺省值填充的方式可以是:按照各个样本特征的缺省值的平均值对缺省值低于第一缺省值阈值的样本特征进行缺省值填充。或者,也可以按照各个样本特征的最高缺省值和最低缺省值之间的任意一个值对缺省值低于第一缺省值阈值的样本特征进行缺省值填充。又或者,还可以按照各个样本特征的最高缺省值和最低缺省值之间的中位缺省值对缺省值低于第一缺省值阈值的样本特征进行缺省值填充等,本实施例对此不作任何限制。

[0115] 接着,可以将处理得到的多个初始正样本和多个初始未知样本映射到同一值空间范围内,得到处于同一值空间范围内的多个正样本和多个未知样本。

[0116] 示例性地,可以按照离差标准化、正态分布密度函数或者非线性函数等任意可行的归一化方式将处理得到的多个初始正样本和多个初始未知样本映射到同一值空间范围。

[0117] 步骤S120,基于多个正样本和多个未知样本,经过至少一个迭代周期的重心聚类,得到与目标服务事件不匹配的多个负样本。

[0118] 在一种可能的实施方式中,针对每个迭代周期,经过该迭代周期内的迭代聚类,得到一个聚类后的正簇和负簇,并在该正簇的多个正样本和该负簇的多个负样本中,计算每个样本与聚类后的正簇的重心之间的第一距离以及与聚类后的负簇的重心之间的第二距离。在此基础上,遍历每个样本,包括每个正样本和每个负样本,若样本计算得到的第一距离小于第二距离,则将该样本作为正样本,若第一距离大于第二距离,则将该样本作为负样本。

[0119] 通过上述经过多个迭代周期分别迭代聚类的方式,可以在每个迭代周期内过滤掉不准确的负样本,从而避免在训练信息推荐模型的过程中错误地将原本与目标服务事件匹配的样本作为负样本的特征信息进行学习而导致的预测误差。

[0120] 另外,在上述重心聚类过程中,在第一个迭代周期中,计算多个正样本的重心,将与计算得到的重心之间的距离小于第一预设阈值的所有样本作为聚类后的第一正簇,将与重心之间的距离大于第二预设阈值的所有样本作为聚类后的第一负簇。在后续每个迭代周期内,都会确定新的正簇和新的负簇。

[0121] 在上述基础上,下面将通过一个具体示例对上述重心聚类过程作进一步说明。

[0122] 请结合参阅图4,为本申请实施例提供的上述步骤S120中进行重心聚类的过程,下面对上述步骤S120包括的各个子步骤进行详细阐述。

[0123] 子步骤S121,将与多个正样本的重心之间的距离小于第一预设阈值的所有样本作为聚类后的第一正簇,将与多个正样本的重心之间的距离大于第二预设阈值的所有样本作为聚类后的第一负簇。

[0124] 子步骤S122,在该第一正簇的多个正样本和该第一负簇的多个负样本中,计算每个样本与该第一正簇的重心之间的第一距离以及与该第一负簇的重心之间的第二距离。

[0125] 子步骤S123,遍历每个样本,若该样本计算得到的第一距离小于第二距离,则将该样本作为正样本,若第一距离大于第二距离,则将该样本作为负样本,以得到对应的第二正簇和第二负簇。

[0126] 子步骤S124,在该第二正簇的多个正样本和该第二负簇的多个负样本中,计算每个样本与该第二正簇的重心之间的第一距离以及与该第二负簇的重心之间的第二距离。

[0127] 子步骤S125,遍历每个样本,若该样本计算得到的第一距离小于第二距离,则将该样本作为正样本,若第一距离大于第二距离,则将该样本作为负样本,以得到对应的第三正簇和第三负簇。

[0128] 子步骤S126,将第三正簇作为新的第二正簇,将第三负簇作为新的第二负簇。

[0129] 子步骤S127,判断是否满足满足迭代停止条件,若不满足,则返回在步骤S124。若满足,则将最后得到的第三负簇中的各个负样本作为与目标服务事件不匹配的负样本。

[0130] 如此,通过重复上述S124-S127的迭代聚类过程,不断更新正簇的重心位置和负簇的重心位置,并不断形成新的正簇和负簇,使得新的正簇中的正样本与目标服务事件的匹配程度越来越大,新的负簇中的负样本与目标服务事件的不匹配程度越来越大,最终在满足迭代停止条件时,将最后得到的第三负簇中的各个负样本作为与目标服务事件不匹配的负样本。

[0131] 其中,上述的迭代停止条件可以包括以下条件中的至少一种:

[0132] 1) 第三正簇中的正样本和第三负簇中的负样本不再发生变化;2) 迭代次数达到设定次数;3) 第三正簇和第三负簇的重心移动距离小于设定距离。

[0133] 其中,在条件1)中,第三正簇中的正样本和第三负簇中的负样本不再发生变化,表明已经形成了最佳的正簇和负簇,可以停止迭代。在条件2)中,为了节省运算量,可以设置迭代次数的最大值,如果迭代次数达到设定次数,可以停止本迭代周期的迭代,将最后得到的各个负样本作为与目标服务事件不匹配的负样本。在条件3)中,如果第三正簇和第三负簇的重心移动距离小于设定距离,说明当前的正簇和负簇已经基本可以满足条件,此时可以停止迭代。

[0134] 需要说明的是,上述迭代停止条件可以结合使用,也可以择一使用,例如,可以在第三正簇中的正样本和第三负簇中的负样本不再发生变化时停止迭代,或者,在迭代次数达到设定次数时停止迭代,或者,在第三正簇和第三负簇的重心移动距离小于设定距离时停止迭代。或者,还可以在迭代次数达到设定次数,并且第三正簇和第三负簇的重心移动距离小于设定距离时,停止迭代。

[0135] 此外,在实际实施过程中,也可以不限于采用上述示例作为迭代停止条件,本领域技术人员可以根据实际需求设计与上述示例不同的迭代停止条件。

[0136] 如此,本实施例通过对训练样本集中的多个正样本和多个未知样本经过至少一个迭代周期的重心聚类,可以得到更为准确的多个与目标服务事件不匹配的负样本,避免在训练信息推荐模型过程中错误地将原本与目标服务事件匹配的样本作为负样本的特征信息进行学习而导致的预测误差。

[0137] 请再次参阅图3,步骤S130,根据多个正样本和多个负样本训练得到信息推荐模

型,并将各个待预测样本输入到训练得到的信息推荐模型中,得到各个待预测样本与目标服务事件的匹配度。

[0138] 在上述方案的基础上,本申请发明人在研究过程中还发现,上述重心聚类只能处理各个样本的线性特征,而无法处理各个样本的非线性特征,因此经过重心聚类后得到的正样本和负样本还会引入一定数量的噪音样本,从而影响后续训练信息推荐模型的准确性。

[0139] 为了解决上述问题,发明人经过研究探索,在上述重心聚类的基础上引入支持向量机(Support Vector Machine,SVM)训练方法,可以处理各个样本的非线性特征,从而进一步地减少噪音样本。

[0140] 例如,在一种可能的实施方式中,可以根据多个正样本和多个负样本训练支持向量机模型,并将训练得到的支持向量机模型作为信息推荐模型。

[0141] 在上述实施方式中,根据多个正样本和多个负样本训练支持向量机模型可以如下方式实现:

[0142] 首先,通过预先设定的核函数对多个正样本和多个负样本进行映射变换。

[0143] 详细地,如果支持向量机通过某非线性变换 $\varphi(x)$,将输入空间映射到高维特征空间。特征空间的维数可能非常高。如果支持向量机的求解只用到内积运算,而在低维输入空间又存在某个函数 $K(x, x')$,它恰好等于在高维空间中这个内积,即 $K(x, x') = \langle \varphi(x), \varphi(x') \rangle$ 。那么支持向量机就不用计算复杂的非线性变换,而由这个函数 $K(x, x')$ 直接得到非线性变换的内积,使大大简化了计算,这样的函数 $K(x, x')$ 称为核函数。

[0144] 在本实施方式中,核函数的选取方式可以是:如果各个正样本和各个负样本线性可分且特征数量和样本数量超过一定数值,则可以选择线性核函数;如果各个正样本和各个负样本线性不可分且特征数量和样本数量低于一定数值,则可以选择RBF(Gaussian Radial Basis Function)核函数;如果各个正样本和各个负样本线性可分且特征数量和样本数量不确定时,可以选择高斯核函数;在最后生成神经网络时,可以选择Sigmoid核函数。

[0145] 接着,将映射变换后的每个正样本和每个负样本输入到支持向量机模型中进行训练,计算映射变换后的每个正样本和每个负样本与超平面之间的距离。而后,根据计算得到的每个正样本和每个负样本与超平面之间的距离计算支持向量机模型的损失函数值,并根据计算得到的损失函数值,利用梯度下降法对支持向量机模型的模型参数和学习率进行调整,接着基于调整后的支持向量机模型返回将映射变换后的每个正样本和每个负样本训练输入到支持向量机模型进行训练,计算映射变换后的每个正样本和每个负样本与超平面之间的距离的步骤,直到满足迭代停止条件,输出训练得到的支持向量机模型。

[0146] 其中,上述迭代停止条件可以是:迭代次数达到设定次数阈值、损失函数值不再下降和/或损失函数值低于设定损失函数值。

[0147] 在上述训练得到支持向量机模型的基础上,在另一种可能的实施方式中,还可以根据训练完成的支持向量机模型对多个正样本和多个负样本进行样本选择,得到多个目标正样本和多个目标负样本。接着,基于多个目标正样本和多个目标负样本对集成树模型进行训练,并将训练得到的集成树模型作为训练后的信息推荐模型。

[0148] 可选地,根据训练完成的支持向量机模型对多个正样本和多个负样本进行样本选择的方式可以是:将每个正样本和每个负样本输入到训练完成的支持向量机模型中,得到

每个正样本和每个负样本与目标服务事件的匹配率。接着,根据每个正样本和每个负样本与目标服务事件的匹配率,将误差值大于第三预设阈值的样本剔除,得到多个目标正样本和多个目标负样本。

[0149] 如此,通过训练完成的支持向量机模型进一步地将各个正样本和负样本中的噪声样本进行剔除,从而进一步地提高每个正样本和每个负样本的准确性,并在此基础上,基于更为准确的多个目标正样本和多个目标负样本对集成树模型进行训练,将训练得到的集成树模型作为训练后的信息推荐模型。

[0150] 在上述基础上,针对每个待预测样本,将该待预测样本输入到训练得到的信息推荐模型中,得到该待预测样本与目标服务事件的匹配度。

[0151] 其中,待预测样本可以理解为当前应用程序实时收集到的服务请求方使用该应用程序生成的服务记录信息的特征信息。例如,仍旧以当前应用程序为网约车应用程序,目标服务事件为买车或者租车为例,网约车应用程序可以实时收集当前注册用户(例如乘客、司机)生成的每个网约车订单的特征信息作为待预测样本,然后将该待预测样本输入到信息推荐模型中,即可得到当前注册用户对买车或者租车的感兴趣程度,也即该待预测样本与目标服务事件的匹配度。

[0152] 步骤S140,针对每个待预测样本,根据该待预测样本与目标服务事件的匹配度确定是否需要该待预测样本的服务请求方终端130推荐与目标服务事件相关的信息。

[0153] 在一种可能的实施方式中,可以判断该待预测样本与目标服务事件的匹配度是否大于第四预设阈值,若该待预测样本与目标服务事件的匹配度大于第四预设阈值,则确定需要向该待预测样本的服务请求方终端130推荐与目标服务事件相关的信息。

[0154] 在上述基础上,若确定需要向该待预测样本的服务请求方终端130推荐与目标服务事件相关的信息,则将该待预测样本添加到训练样本中的多个正样本中。

[0155] 此外,有时候尽管已经确定该待预测样本与目标服务事件的匹配度不大于第四预设阈值,但是以此确定不需要向该待预测样本的服务请求方终端130推荐与目标服务事件相关的信息可能还是不太准确,故在本实施例中,还需要进一步判断该待预测样本是否大于第五预设阈值。也即,虽然该待预测样本与目标服务事件的匹配度不大于第四预设阈值,但是如果该匹配度仍旧大于第五预设阈值,则向待预测样本的服务请求方终端130发送提示信息以提示该服务请求方终端130的用户选择是否接收目标服务事件相关的信息。若接收到该服务请求方终端130发送的不接收目标服务事件相关的信息的第一指示信息,则确定不需要向该待预测样本推荐与目标服务事件相关的信息,并将该待预测样本添加到与目标服务事件不匹配的多个负样本中。若接收到该服务请求方终端130发送的接收目标服务事件相关的信息的第二指示信息,则确定需要向该待预测样本推荐与目标服务事件相关的信息,并将该待预测样本添加到训练样本中的多个正样本中。

[0156] 若该待预测样本与目标服务事件的匹配度不大于第五预设阈值,则可以确定不需要向该待预测样本推荐与目标服务事件相关的信息,并将该待预测样本添加到与目标服务事件不匹配的多个负样本中。

[0157] 基于上述设计,不仅可以该待预测样本与目标服务事件的匹配度确定是否需要该待预测样本的服务请求方终端130推荐与目标服务事件相关的信息,同时还可以根据确定结果将该待预测样本重新作为正样本或者负样本,从而增加正样本或者负样本的数

量,并返回上述训练步骤,不断迭代训练,从而不断提高信息推荐模型的预测准确性,为服务请求方提供更多便利、高效的目标服务。

[0158] 在进一步的实施方式中,请参阅图5,在该步骤S140之后,该信息推荐方法还可以包括如下步骤:

[0159] 步骤S150,若确定需要向该待预测样本的服务请求方终端130推荐与目标服务事件相关的信息,则向各个提供目标服务事件的服务的第三方服务器发送信息获取请求。

[0160] 步骤S150,接收各个第三方服务器响应信息获取请求发送的与目标服务事件相关的信息,并将与目标服务事件相关的信息推荐给该待预测样本的服务请求方终端130。

[0161] 仍旧以目标服务事件为买车或者租车为例,如果确定需要向该待预测样本的服务请求方终端130推荐与买车或者租车相关的信息,则向各个提供买车或者租车服务的第三方服务器(例如卖好车网、汽车之家网、瓜子二手车直卖网、一嗨租车网、神州租车网等的第三方服务器)发送信息获取请求。

[0162] 可选地,该信息获取请求中可包括该待预测样本的特征信息,例如司机或者乘客的行为习惯、偏好、驾龄、年龄、性别等信息。接着,各个提供买车或者租车服务的第三方服务器响应信息获取请求后获取与该待预测样本的特征信息匹配的买车或者租车相关的信息,然后返回上述信息给该服务器110。该服务器110在接收到各个第三方服务器返回的买车或者租车相关的信息,将买车或者租车相关的信息推荐给该待预测样本的服务请求方终端130。

[0163] 图6示出了本申请一些实施例提供的信息推荐装置300的功能模块图,该信息推荐装置300实现的功能可以对应上述方法执行的步骤。该信息推荐装置300可以理解为上述服务器110,或服务器110的处理器,也可以理解为独立于上述服务器110或处理器之外的在服务器110控制下实现本申请功能的组件,如图6所示,该信息推荐装置300可以包括获得模块310、重心聚类模块320、训练模块330以及确定模块340,下面分别该信息推荐装置300的各个功能模块的功能进行详细阐述。

[0164] 获得模块310,可以用于获得训练样本集,训练样本集包括多个正样本和多个未知样本。可以理解,该获得模块310可以用于执行上述步骤S110,关于该获得模块310的详细实现方式可以参照上述对步骤S110有关的内容。

[0165] 重心聚类模块320,可以用于基于多个正样本和多个未知样本,经过至少一个迭代周期的重心聚类,得到与目标服务事件不匹配的多个负样本,其中,每个正样本与目标服务事件匹配。可以理解,该重心聚类模块320可以用于执行上述步骤S120,关于该重心聚类模块320的详细实现方式可以参照上述对步骤S120有关的内容。

[0166] 训练模块330,可以用于根据多个正样本和多个负样本训练得到信息推荐模型,并将各个待预测样本输入到训练得到的信息推荐模型中,得到各个待预测样本与目标服务事件的匹配度。可以理解,该训练模块330可以用于执行上述步骤S130,关于该训练模块330的详细实现方式可以参照上述对步骤S130有关的内容。

[0167] 确定模块340,可以用于针对每个待预测样本,根据该待预测样本与目标服务事件的匹配度确定是否需要向该待预测样本的服务请求方终端130推荐与目标服务事件相关的信息。可以理解,该确定模块340可以用于执行上述步骤S140,关于该确定模块340的详细实现方式可以参照上述对步骤S140有关的内容。

[0168] 在一种可能的实施方式中,获得模块310,具体可以通过以下方式获得多个正样本和多个未知样本:

[0169] 获得多个初始正样本和多个初始未知样本;

[0170] 对每个初始正样本和每个初始未知样本的样本特征进行归一化处理,得到处于同一值空间范围内的多个正样本和多个未知样本。

[0171] 在一种可能的实施方式中,获得模块310,具体可以通过以下方式获得多个初始正样本和多个初始未知样本:

[0172] 获得与目标服务事件匹配的多个第一用户的第一服务记录信息以及第一用户之外的多个第二用户的第二服务记录信息;

[0173] 从第一服务记录信息中提取特征信息,以得到多个初始正样本,并从第二服务记录信息中提取特征信息,以得到多个初始未知样本。

[0174] 在一种可能的实施方式中,获得模块310,具体可以通过以下方式得到处于同一值空间范围内的多个正样本和多个未知样本:

[0175] 获取每个初始正样本和每个初始未知样本的各个样本特征的缺省值;

[0176] 对缺省值低于第一缺省值阈值的样本特征进行缺省值填充,并剔除缺省值大于第二缺省值阈值的样本特征,得到处理后的多个初始正样本和多个初始未知样本;

[0177] 将处理得到的多个初始正样本和多个初始未知样本映射到同一值空间范围内,得到处于同一值空间范围内的多个正样本和多个未知样本。

[0178] 在一种可能的实施方式中,重心聚类模块320,具体用于可以通过以下方式得到与目标服务事件不匹配的多个负样本的:

[0179] 针对每个迭代周期,经过该迭代周期内的迭代聚类,得到一个聚类后的正簇和负簇;

[0180] 在该正簇的多个正样本和该负簇的多个负样本中,计算每个样本与聚类后的正簇的重心之间的第一距离以及与聚类后的负簇的重心之间的第二距离;

[0181] 遍历每个样本,若样本计算得到的第一距离小于第二距离,则将该样本作为正样本,若第一距离大于第二距离,则将该样本作为负样本。

[0182] 在一种可能的实施方式中,重心聚类模块320,具体用于可以通过以下方式执行第一个迭代周期的重心聚类:

[0183] 在第一个迭代周期中,计算多个正样本的重心;

[0184] 将与计算得到的重心之间的距离小于第一预设阈值的所有样本作为聚类后的第一正簇,将与重心之间的距离大于第二预设阈值的所有样本作为聚类后的第一负簇。

[0185] 在一种可能的实施方式中,重心聚类模块320,具体用于可以通过以下方式得到与目标服务事件不匹配的多个负样本的步骤,包括:

[0186] 将与多个正样本的重心之间的距离小于第一预设阈值的所有样本作为聚类后的第一正簇,将与多个正样本的重心之间的距离大于第二预设阈值的所有样本作为聚类后的第一负簇;

[0187] 在该第一正簇的多个正样本和该第一负簇的多个负样本中,计算每个样本与该第一正簇的重心之间的第一距离以及与该第一负簇的重心之间的第二距离;

[0188] 遍历每个样本,若该样本计算得到的第一距离小于第二距离,则将该样本作为正

样本,若第一距离大于第二距离,则将该样本作为负样本,以得到对应的第二正簇和第二负簇;

[0189] 在该第二正簇的多个正样本和该第二负簇的多个负样本中,计算每个样本与该第二正簇的重心之间的第一距离以及与该第二负簇的重心之间的第二距离;

[0190] 遍历每个样本,若该样本计算得到的第一距离小于第二距离,则将该样本作为正样本,若第一距离大于第二距离,则将该样本作为负样本,以得到对应的第三正簇和第三负簇;

[0191] 将第三正簇作为新的第二正簇,将第三负簇作为新的第二负簇,并返回在该第二正簇的多个正样本和该第二负簇的多个负样本中,计算每个样本与该第二正簇的重心之间的第一距离以及与该第二负簇的重心之间的第二距离的步骤,直到满足迭代停止条件,将最后得到的第三负簇中的各个负样本作为与目标服务事件不匹配的负样本。

[0192] 在一种可能的实施方式中,迭代停止条件包括以下条件中的至少一种:

[0193] 第三正簇中的正样本和第三负簇中的负样本不再发生变化;

[0194] 迭代次数达到设定次数;

[0195] 第三正簇和第三负簇的重心移动距离小于设定距离。

[0196] 在一种可能的实施方式中,训练模块330,具体可以通过以下方式训练得到信息推荐模型:

[0197] 根据多个正样本和多个负样本训练支持向量机模型,并将训练得到的支持向量机模型作为信息推荐模型。

[0198] 在一种可能的实施方式中,训练模块330,具体可以通过以下方式训练得到信息推荐模型:

[0199] 根据多个正样本和多个负样本训练支持向量机模型;

[0200] 根据训练完成的支持向量机模型对多个正样本和多个负样本进行样本选择,得到多个目标正样本和多个目标负样本;

[0201] 基于多个目标正样本和多个目标负样本对集成树模型进行训练,并将训练得到的集成树模型作为训练后的信息推荐模型。

[0202] 在一种可能的实施方式中,训练模块330,具体可以通过以下方式训练支持向量机模型:

[0203] 通过预先设定的核函数对多个正样本和多个负样本进行映射变换;

[0204] 将映射变换后的每个正样本和每个负样本输入到支持向量机模型中进行训练,计算映射变换后的每个正样本和每个负样本与超平面之间的距离;

[0205] 根据计算得到的每个正样本和每个负样本与超平面之间的距离计算支持向量机模型的损失函数值;

[0206] 根据计算得到的损失函数值,利用梯度下降法对支持向量机模型的模型参数和学习率进行调整,并基于调整后的支持向量机模型返回将映射变换后的每个正样本和每个负样本训练输入到支持向量机模型进行训练,计算映射变换后的每个正样本和每个负样本与超平面之间的距离的步骤,直到满足迭代停止条件,输出训练得到的支持向量机模型。

[0207] 在一种可能的实施方式中,训练模块330,具体可以通过以下方式得到多个目标正样本和多个目标负样本:

[0208] 将每个正样本和每个负样本输入到训练完成的支持向量机模型中,得到每个正样本和每个负样本与目标服务事件的匹配率;

[0209] 根据每个正样本和每个负样本与目标服务事件的匹配率将误差值大于第三预设阈值的样本剔除,得到多个目标正样本和多个目标负样本。

[0210] 在一种可能的实施方式中,确定模块340可以通过以下方式确定是否需要向该待预测样本的服务请求方终端130推荐与目标服务事件相关的信息:

[0211] 判断该待预测样本与目标服务事件的匹配度是否大于第四预设阈值;

[0212] 若该待预测样本与目标服务事件的匹配度大于第四预设阈值,则确定需要向该待预测样本的服务请求方终端130推荐与目标服务事件相关的信息。

[0213] 在一种可能的实施方式中,若确定需要向该待预测样本的服务请求方终端130推荐与目标服务事件相关的信息,确定模块340,具体还用于:

[0214] 将该待预测样本添加到训练样本中的多个正样本中。

[0215] 在一种可能的实施方式中,若该待预测样本与目标服务事件的匹配度大于第五预设阈值且不大于第四预设阈值,则确定模块340具体还可以用于:

[0216] 向待预测样本的服务请求方终端130发送提示信息以提示该服务请求方终端130的用户选择是否接收目标服务事件相关的信息;

[0217] 若接收到该服务请求方终端130发送的不接收目标服务事件相关的信息的第一指示信息,则确定不需要向该待预测样本推荐与目标服务事件相关的信息,并将该待预测样本添加到与目标服务事件不匹配的多个负样本中;

[0218] 若接收到该服务请求方终端130发送的接收目标服务事件相关的信息的第二指示信息,则确定需要向该待预测样本推荐与目标服务事件相关的信息,并将该待预测样本添加到训练样本中的多个正样本中;以及

[0219] 若该待预测样本与目标服务事件的匹配度不大于第五预设阈值,则确定模块340,具体还用于确定不需要向该待预测样本推荐与目标服务事件相关的信息,并将该待预测样本添加到与目标服务事件不匹配的多个负样本中。

[0220] 在一种可能的实施方式中,请进一步地参阅图7,若确定需要向该待预测样本的服务请求方终端130推荐与目标服务事件相关的信息,信息推荐装置还可以包括信息推荐模块350,信息推荐模块350可以用于向各个提供目标服务事件的服务的第三方服务器发送信息获取请求,接收各个第三方服务器响应信息获取请求发送的与目标服务事件相关的信息,并将与目标服务事件相关的信息推荐给该待预测样本的服务请求方终端130。

[0221] 上述模块可以经由有线连接或无线连接彼此连接或通信。有线连接可以包括金属线缆、光缆、混合线缆等,或其任意组合。无线连接可以包括通过LAN、WAN、蓝牙、ZigBee、或NFC等形式的连接,或其任意组合。两个或更多个模块可以组合为单个模块,并且任何一个模块可以分成两个或更多个单元。

[0222] 所属领域的技术人员可以清楚地了解到,为描述的方便和简洁,上述描述的系统 and 装置的具体工作过程,可以参考方法实施例中的对应过程,本申请中不再赘述。在本申请所提供的几个实施例中,应该理解到,所揭露的系统、装置和方法,可以通过其它的方式实现。以上所描述的装置实施例仅仅是示意性的,例如,所述模块的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,又例如,多个模块或组件可以结合或者可以集

成到另一个系统,或一些特征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些通信接口,装置或模块的间接耦合或通信连接,可以是电性,机械或其它的形式。

[0223] 所述作为分离部件说明的模块可以是或者也可以不是物理上分开的,作为模块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单元来实现本实施例方案的目的。

[0224] 另外,在本申请各个实施例中的各功能单元可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中。

[0225] 所述功能如果以软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个处理器可执行的非易失的计算机可读取存储介质中。基于这样的理解,本申请的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案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一个存储介质中,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)执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储介质包括:U盘、移动硬盘、ROM、RAM、磁碟或者光盘等各种可以存储程序代码的介质。

[0226] 以上仅为本申请的具体实施方式,但本申请的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本申请揭露的技术范围内,可轻易想到变化或替换,都应涵盖在本申请的保护范围之内。因此,本申请的保护范围应以权利要求的保护范围为准。

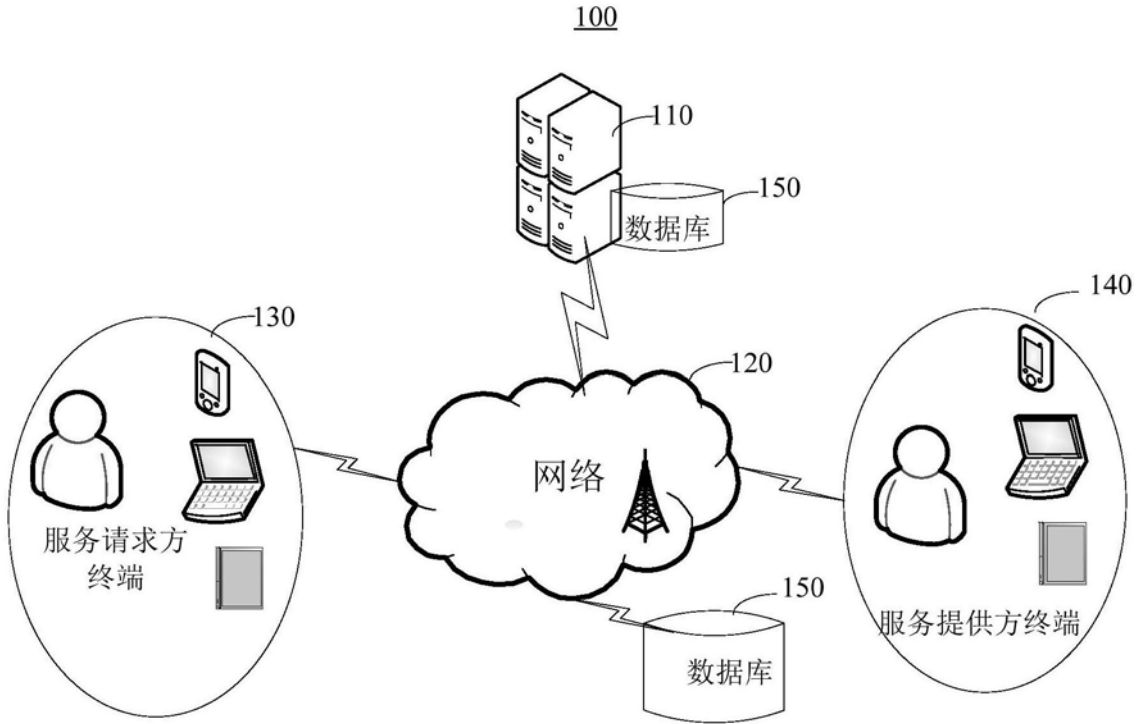


图1

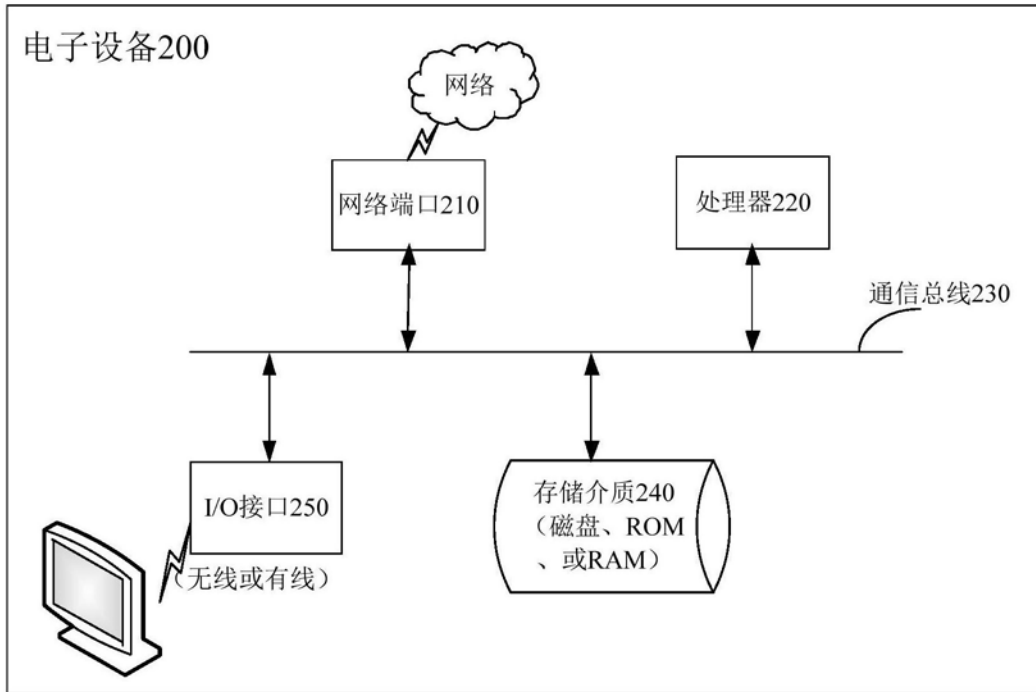


图2

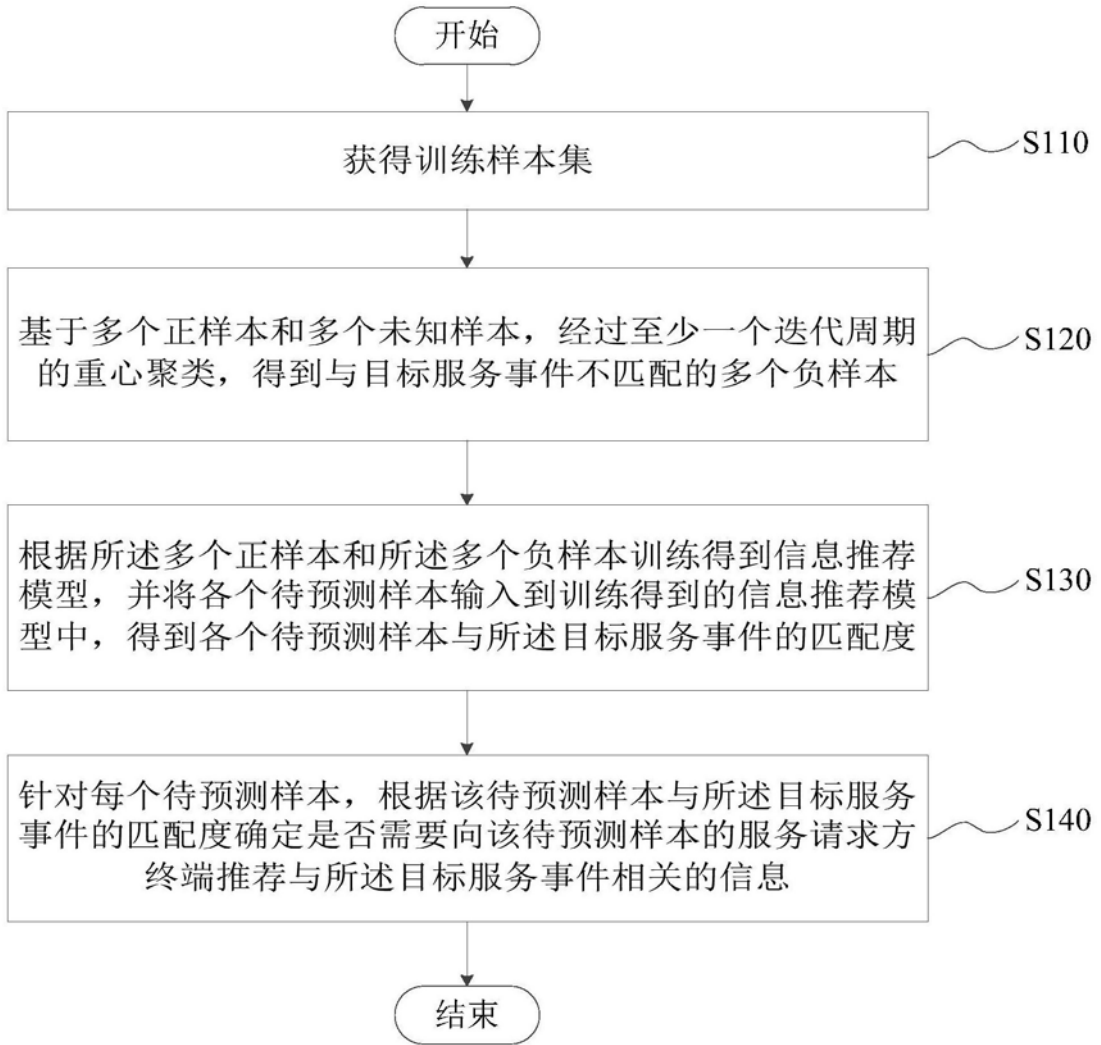


图3

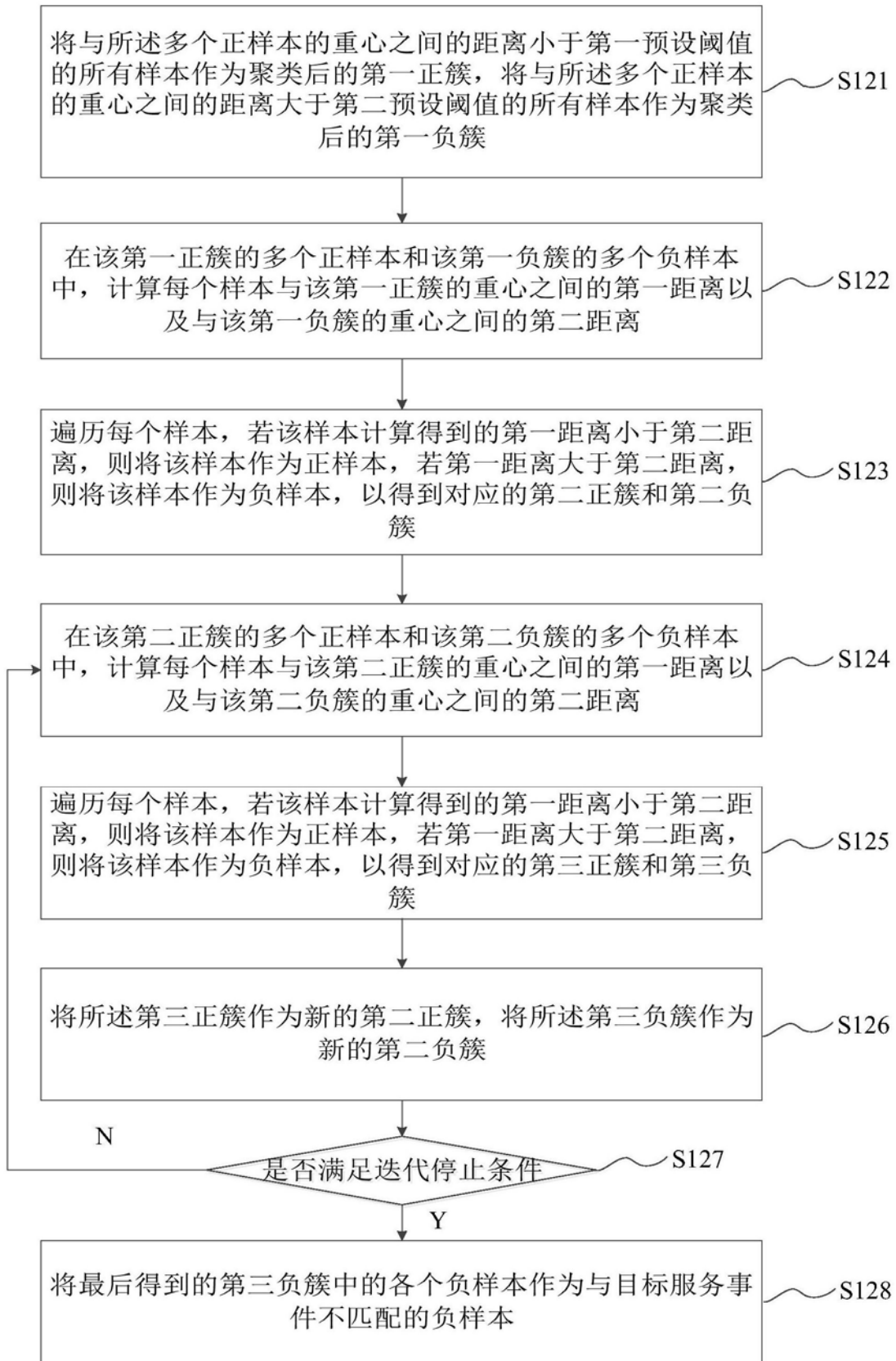


图4

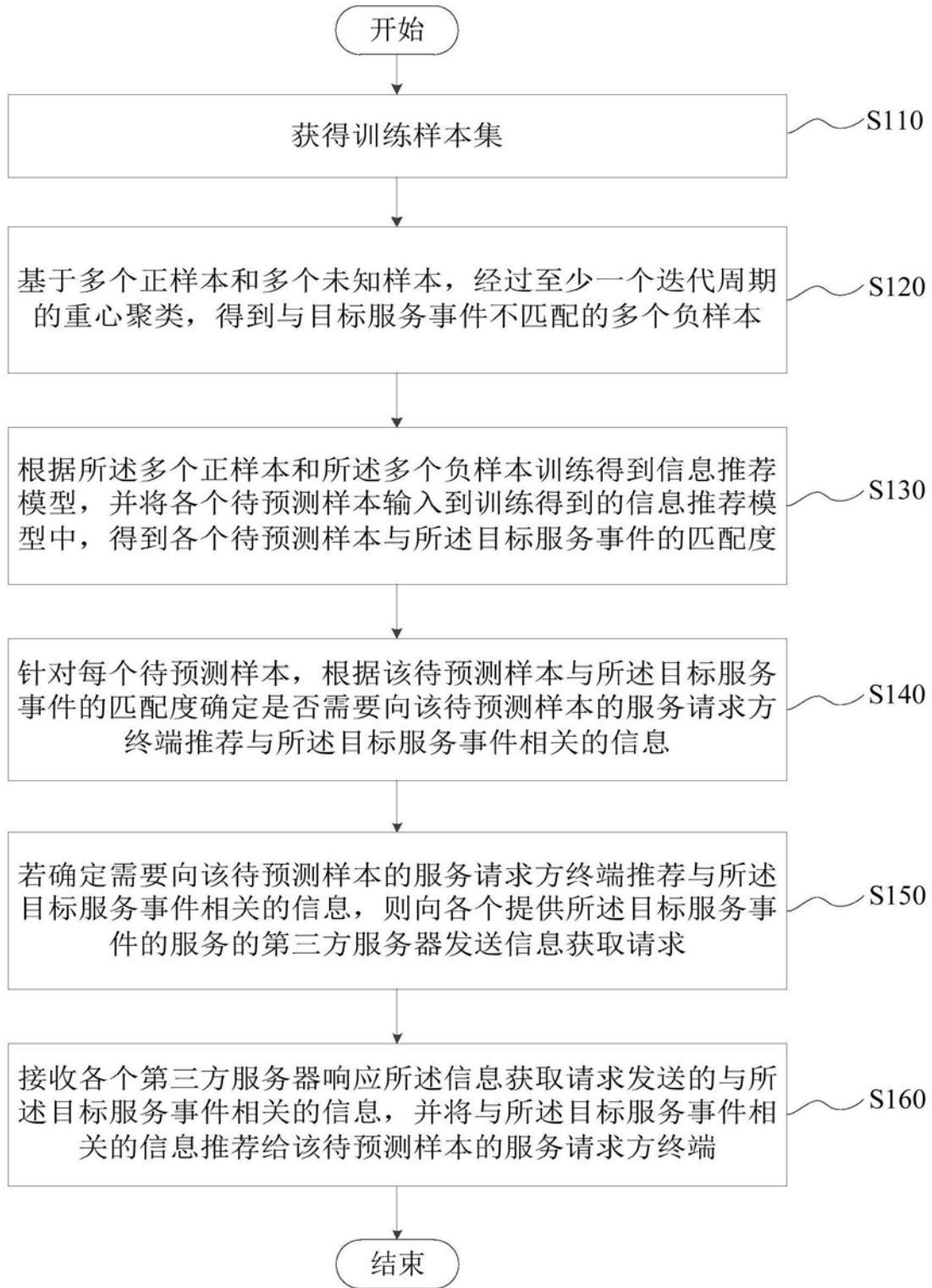


图5

300

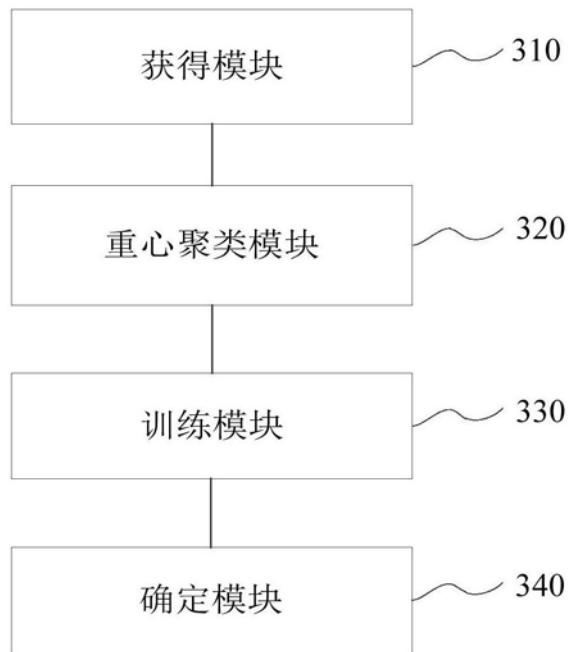


图6

300

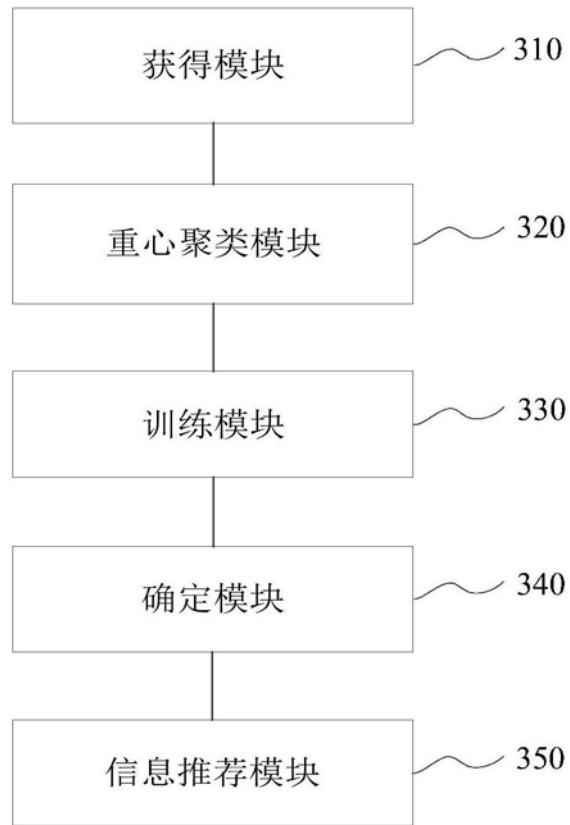


图7