



(12) 发明专利

(10) 授权公告号 CN 113610552 B

(45) 授权公告日 2024.07.19

(21) 申请号 202110713356.X

G06N 3/045 (2023.01)

(22) 申请日 2021.06.25

G06N 3/042 (2023.01)

(65) 同一申请的已公布的文献号

G06N 3/0464 (2023.01)

申请公布号 CN 113610552 A

G06N 3/084 (2023.01)

(43) 申请公布日 2021.11.05

G06N 3/0985 (2023.01)

(73) 专利权人 清华大学

G06N 5/045 (2023.01)

地址 100084 北京市海淀区双清路30号清华大学

(56) 对比文件

Guozhen Zhang等.A Counterfactual Modeling Framework for Churn Prediction.《ACM》.2022,全文.

(72) 发明人 李勇 张国祯 金德鹏

审查员 宿渊源

(74) 专利代理机构 北京路浩知识产权代理有限公司 11002

专利代理师 王治东

(51) Int. Cl.

G06Q 30/0202 (2023.01)

G06Q 30/0201 (2023.01)

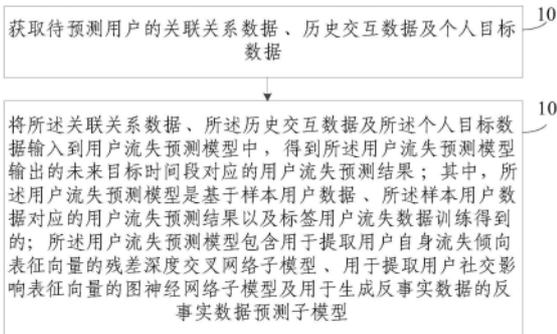
权利要求书2页 说明书13页 附图4页

(54) 发明名称

一种用户流失预测方法及装置

(57) 摘要

本发明提供一种用户流失预测方法及装置。该方法包括:获取待预测用户的关联关系数据、历史交互数据及个人目标数据;将关联关系数据、历史交互数据及个人目标数据输入到用户流失预测模型中,得到用户流失预测模型输出的未来目标时间段对应的用户流失预测结果;用户流失预测模型是基于样本用户数据、样本用户数据对应的用户流失预测结果及标签用户流失数据训练得到的;用户流失预测模型包含用于提取用户自身流失倾向表征向量的残差深度交叉网络子模型、用于提取用户社交影响表征向量的图神经网络子模型及用于生成反事实数据的反事实数据预测子模型。采用本发明方法,基于反事实推理建模架构引入用户间相互关系因素,提高了用户流失预测的准确度。



1. 一种用户流失预测方法,其特征在于,包括:

获取待预测用户的关联关系数据、历史交互数据及个人目标数据;

将所述关联关系数据、所述历史交互数据及所述个人目标数据输入到用户流失预测模型中,得到所述用户流失预测模型输出的未来目标时间段对应的用户流失预测结果;

其中,所述用户流失预测模型是基于样本用户数据、所述样本用户数据对应的用户流失预测结果以及标签用户流失数据训练得到的;

所述用户流失预测模型包含用于提取用户自身流失倾向表征向量的残差深度交叉网络子模型、用于提取用户社交影响表征向量的图神经网络子模型及用于生成反事实数据的反事实数据预测子模型;

所述图神经网络子模型包括基于社交驱动的图卷积组件和边学习组件;其中,所述边学习组件,用于将输入的节点特征、边特征以及节点间交互特征进行拼接,并利用预设线性网络模型学习得到针对关联关系网络中每一条边的蒙版向量;所述节点特征、所述边特征以及所述节点间交互特征是分别基于对所述个人目标数据、所述关联关系数据以及所述历史交互数据进行建模及特征提取得到的;所述图卷积组件,用于针对所述关联关系网络中的目标节点,基于所述边学习组件得到的对应蒙版向量以及相邻节点的状态向量,通过注意力机制更新所述目标节点的状态向量;将所述目标节点的状态向量作为所述用户社交影响表征向量。

2. 根据权利要求1所述的用户流失预测方法,其特征在于,还包括:预先获取样本用户数据,并基于所述样本用户数据对初始用户流失预测模型进行训练,得到相应的所述用户流失预测模型;

所述基于所述样本用户数据对初始用户流失预测模型进行训练,得到相应的所述用户流失预测模型,具体包括:

将所述样本用户数据输入到初始用户流失预测模型中,利用残差深度交叉网络子模型从样本用户数据中提取出用户自身流失倾向表征向量,利用社交驱动的图神经网络子模型从样本用户数据中提取出用户社交影响表征向量;利用反事实数据预测子模型,基于相应的观测数据生成反事实样本数据;

将所述用户自身流失倾向表征向量、所述用户社交影响表征向量输入到因果预测模型中,利用所述反事实样本数据和相应的事实数据共同训练初始用户流失预测模型,得到建模社交影响因果效应的用户流失预测模型。

3. 根据权利要求2所述的用户流失预测方法,其特征在于,还包括:

在将所述样本用户数据输入到初始用户流失预测模型中训练之前,预先对获取的原始样本用户数据中的离散变量进行独热编码,对所述原始样本用户数据中的连续变量进行归一化处理;

将独热编码得到的编码结果与归一化处理得到的处理结果进行拼接后,利用全连接神经网络模型将初始样本用户数据特征映射到连续向量空间,得到连续特征向量表示的样本用户数据。

4. 根据权利要求1所述的用户流失预测方法,其特征在于,所述残差深度交叉网络子模型包含深度交叉网络,所述深度交叉网络包含深度残差网络;其中,所述深度残差网络,用于基于第1层交叉网络的隐节点向量和第1层深度网络的隐节点向量,得到第1+1层交叉网

网络的隐节点向量和第1+1层深度网络的隐节点向量；并基于所述第1+1层交叉网络的隐节点向量和所述第1+1层深度网络的隐节点向量，得到所述用户自身流失倾向表征向量。

5. 根据权利要求1所述的用户流失预测方法，其特征在于，所述反事实数据预测子模型包括反事实数据扩增模块和反事实数据预测模块；

其中，所述反事实数据扩增模块，用于基于预设的因果关系模型以及反事实数据扩增推论规则，得到反事实数据；

所述反事实数据预测模块，用于通过预设的三头预测的多任务预测器对所述反事实数据、相应的事实数据以及处理变量进行预测，基于因果正则项建模社交影响因果效应，以得到对应的用户流失预测结果。

6. 根据权利要求1所述的用户流失预测方法，其特征在于，所述个人目标数据包含用户的基础属性数据和历史行为数据；

所述关联关系数据是基于存在用户交互行为的应用程序对应的全量交互历史数据构建的关联关系网络得到的。

7. 一种用户流失预测装置，其特征在于，包括：

用户数据获取单元，用于获取待预测用户的关联关系数据、历史交互数据及个人目标数据；

用户流失预测单元，用于将所述关联关系数据、所述历史交互数据及所述个人目标数据输入到用户流失预测模型中，得到所述用户流失预测模型输出的未来目标时间段对应的用户流失预测结果；

其中，所述用户流失预测模型是基于样本用户数据、所述样本用户数据对应的用户流失预测结果以及标签用户流失数据训练得到的；

所述用户流失预测模型包含用于提取用户自身流失倾向表征向量的残差深度交叉网络子模型、用于提取用户社交影响表征向量的图神经网络子模型及用于生成反事实数据的反事实数据预测子模型；

所述图神经网络子模型包括基于社交驱动的图卷积组件和边学习组件；其中，所述边学习组件，用于将输入的节点特征、边特征以及节点间交互特征进行拼接，并利用预设线性网络模型学习得到针对关联关系网络中每一条边的蒙版向量；所述节点特征、所述边特征以及所述节点间交互特征是分别基于对所述个人目标数据、所述关联关系数据以及所述历史交互数据进行建模及特征提取得到的；所述图卷积组件，用于针对所述关联关系网络中的目标节点，基于所述边学习组件得到的对应蒙版向量以及相邻节点的状态向量，通过注意力机制更新所述目标节点的状态向量；将所述目标节点的状态向量作为所述用户社交影响表征向量。

8. 一种电子设备，包括存储器、处理器以及存储在存储器上并可在处理器上运行的计算机程序，其特征在于，所述处理器执行所述程序时实现如权利要求1-6任意一项所述的用户流失预测方法的步骤。

9. 一种处理器可读存储介质，所述处理器可读存储介质上存储有计算机程序，其特征在于，该计算机程序被处理器执行时实现如权利要求1-6任意一项所述的用户流失预测方法的步骤。

一种用户流失预测方法及装置

技术领域

[0001] 本发明涉及大数据分析技术领域,具体涉及一种用户流失预测方法和装置。另外,还涉及一种电子设备及处理器可读存储介质。

背景技术

[0002] 近年来,随着计算机技术的快速发展,基于社交网络或者存在用户间交互关系的应用平台越来越多,人们对平台用户的留存情况也越来越重视,因此如何快速、有效对平台用户的流失情况进行预测成为本领域技术人员研究的重点。尤其是对于现代互联网公司来说,在市场竞争白热化的阶段,精准地预测用户流失并采取有针对性的干预措施已经成为公司存活和盈利的关键。

[0003] 用户流失预测(User Churn Prediction)是指预测某个用户未来一段时间是否会停止使用某种服务或应用。目前,绝大多数的用户流失预测技术都只考虑使用用户自身的流失原因。预测方法通常是利用用户属性和行为属性数据构建特征,并将这些构建好的特征输入诸如简单贝叶斯模型、支持向量机(Support Vector Machine,SVM)或随机森林(Random Forest,RF)等模型来预测。然而,这类方法的核心问题是缺乏对用户之间相互关系的建模。尤其是在社交网络高度发达的时代,用户之间的相互联系日益紧密,社交影响对用户流失存在显著的效应。因此,缺乏对社交影响的建模导致用户流失预测模型准确性差,不适用于大多数现有的互联网应用。

[0004] 为了解决上述问题,现有技术中采用的建模社交影响的用户流失预测模型主要可以分为两类:第一类是利用社交网络或者用户交互网络的特征来建模社交影响的效应;第二类是将用户的流失建模为一个扩散过程,认为一个用户如果在其朋友流失之后相继流失了,那么流失就发生了扩散;之后,利用传播模型来模拟这个流失扩散过程并给出对未来的预测。然而,这两类方法均没有抓住社交影响的因果本质。一个用户是否受到其关联关系(比如社交关系)用户的影响是一个因果的问题,用户的相继流失并不能代表关联关系影响的存在。这导致现有技术中采用的方法一方面准确性较差,另一方面对预测结果的可解释性也较差。因此,如何有效提高当前用户流失预测的准确度和可解释性成为亟待解决的技术问题。

发明内容

[0005] 为此,本发明提供一种用户流失预测方法及装置,以解决现有技术中存在的用户流失预测方案中预测准确度和可解释性较差的问题。

[0006] 第一方面,本发明提供一种用户流失预测方法,其特征在于,包括:获取待预测用户的关联关系数据、历史交互数据及个人目标数据;

[0007] 将所述关联关系数据、所述历史交互数据及所述个人目标数据输入到用户流失预测模型中,得到所述用户流失预测模型输出的未来目标时间段对应的用户流失预测结果;

[0008] 其中,所述用户流失预测模型是基于样本用户数据、所述样本用户数据对应的用

户流失预测结果以及标签用户流失数据训练得到的；

[0009] 所述用户流失预测模型包含用于提取用户自身流失倾向表征向量的残差深度交叉网络子模型、用于提取用户社交影响表征向量的图神经网络子模型及用于生成反事实数据的反事实数据预测子模型。

[0010] 进一步的,所述残差深度交叉网络子模型包含深度交叉网络,所述深度交叉网络包含深度残差网络;其中,所述深度残差网络,用于基于第1层交叉网络的隐节点向量和第1层深度网络的隐节点向量,得到第1+1层交叉网络的隐节点向量和第1+1层深度网络的隐节点向量;并基于所述第1+1层交叉网络的隐节点向量和所述第1+1层深度网络的隐节点向量,得到所述用户自身流失倾向表征向量。

[0011] 进一步的,所述图神经网络子模型包括基于社交驱动的图卷积组件和边学习组件;

[0012] 其中,所述边学习组件,用于将输入的节点特征、边特征以及节点间交互特征进行拼接,并利用预设线性网络模型学习得到针对关联关系网络中每一条边的蒙版向量;所述节点特征、所述边特征以及所述节点间交互特征是分别基于对所述个人目标数据、所述关联关系数据以及所述历史交互数据进行建模及特征提取得到的;

[0013] 所述图卷积组件,用于针对所述关联关系网络中的目标节点,基于所述边学习组件得到的对应蒙版向量以及相邻节点的状态向量,通过注意力机制更新所述目标节点的状态向量;将所述目标节点的状态向量作为所述用户社交影响表征向量。

[0014] 进一步的,所述反事实数据预测子模型包括反事实数据扩增模块和反事实数据预测模块;

[0015] 其中,所述反事实数据扩增模块,用于基于预设的因果关系模型以及反事实数据扩增推论规则,得到反事实数据;

[0016] 所述反事实数据预测模块,用于通过预设的三头预测的多任务预测器对所述反事实数据、相应的事实数据以及处理变量进行预测,基于因果正则项建模社交影响因果效应,以得到对应的用户流失预测结果。

[0017] 进一步的,所述的用户流失预测方法,还包括:预先获取样本用户数据,并基于所述样本用户数据对初始用户流失预测模型进行训练,得到相应的所述用户流失预测模型;

[0018] 所述基于所述样本用户数据对初始用户流失预测模型进行训练,得到相应的所述用户流失预测模型,具体包括:

[0019] 将所述样本用户数据输入到初始用户流失预测模型中,利用残差深度交叉网络子模型从样本用户数据中提取出用户自身流失倾向表征向量,利用社交驱动的图神经网络子模型从样本用户数据中提取出用户社交影响表征向量;

[0020] 利用反事实数据预测子模型,基于相应的观测数据生成反事实样本数据;

[0021] 将所述用户自身流失倾向表征向量、所述用户社交影响表征向量输入到因果预测模型中,利用所述反事实样本数据和相应的事实数据共同训练初始用户流失预测模型,得到建模社交影响因果效应的用户流失预测模型。

[0022] 进一步的,所述的用户流失预测方法,还包括:

[0023] 在将所述样本用户数据输入到初始用户流失预测模型中训练之前,预先对获取的原始样本用户数据中的离散变量进行独热编码,对所述原始样本用户数据中的连续变量进

行归一化处理；

[0024] 将独热编码得到的编码结果与归一化处理得到的处理结果进行拼接后,利用全连接神经网络模型将初始样本用户数据特征映射到连续向量空间,得到连续特征向量表示的样本用户数据。

[0025] 进一步的,所述个人目标数据包含用户的基础属性数据和历史行为数据;所述关联关系数据是基于存在用户交互行为的应用程序对应的全量交互历史数据构建的关联关系网络得到的。

[0026] 第二方面,本发明提供一种用户流失预测装置,包括:

[0027] 用户数据获取单元,用于获取待预测用户的关联关系数据、历史交互数据及个人目标数据;

[0028] 用户流失预测单元,用于将所述关联关系数据、所述历史交互数据及所述个人目标数据输入到用户流失预测模型中,得到所述用户流失预测模型输出的未来目标时间段对应的用户流失预测结果;

[0029] 其中,所述用户流失预测模型是基于样本用户数据、所述样本用户数据对应的用户流失预测结果以及标签用户流失数据训练得到的;

[0030] 所述用户流失预测模型包含用于提取用户自身流失倾向表征向量的残差深度交叉网络子模型、用于提取用户社交影响表征向量的图神经网络子模型及用于生成反事实数据的反事实数据预测子模型。

[0031] 进一步的,所述残差深度交叉网络子模型包含深度交叉网络,所述深度交叉网络包含深度残差网络;其中,所述深度残差网络,用于基于第1层交叉网络的隐节点向量和第1层深度网络的隐节点向量,得到第1+1层交叉网络的隐节点向量和第1+1层深度网络的隐节点向量;并基于所述第1+1层交叉网络的隐节点向量和所述第1+1层深度网络的隐节点向量,得到所述用户自身流失倾向表征向量。

[0032] 进一步的,所述图神经网络子模型包括基于社交驱动的图卷积组件和边学习组件;

[0033] 其中,所述边学习组件,用于将输入的节点特征、边特征以及节点间交互特征进行拼接,并利用预设线性网络模型学习得到针对关联关系网络中每一条边的蒙版向量;所述节点特征、所述边特征以及所述节点间交互特征是分别基于对个人目标数据、所述关联关系数据以及所述历史交互数据进行建模及特征提取得到的;

[0034] 所述图卷积组件,用于针对所述关联关系网络中的目标节点,基于所述边学习组件得到的对应蒙版向量以及相邻节点的状态向量,通过注意力机制更新所述目标节点的状态向量;将所述目标节点的状态向量作为所述用户社交影响表征向量。

[0035] 进一步的,所述反事实数据预测子模型包括反事实数据扩增模块和反事实数据预测模块;

[0036] 其中,所述反事实数据扩增模块,用于基于预设的因果关系模型以及反事实数据扩增推论规则,得到反事实数据;

[0037] 所述反事实数据预测模块,用于通过预设的三头预测的多任务预测器对所述反事实数据、相应的事实数据以及处理变量进行预测,基于因果正则项建模社交影响因果效应,以得到对应的用户流失预测结果。

[0038] 进一步的,所述的用户流失预测装置,还包括:模型训练单元,用于预先获取样本用户数据,并基于所述样本用户数据对初始用户流失预测模型进行训练,得到相应的所述用户流失预测模型;

[0039] 所述基于所述样本用户数据对初始用户流失预测模型进行训练,得到相应的所述用户流失预测模型,具体包括:

[0040] 将所述样本用户数据输入到初始用户流失预测模型中,利用残差深度交叉网络子模型从样本用户数据中提取出用户自身流失倾向表征向量,利用社交驱动的图神经网络子模型从样本用户数据中提取出用户社交影响表征向量;

[0041] 利用反事实数据预测子模型,基于相应的观测数据生成反事实样本数据;

[0042] 将所述用户自身流失倾向表征向量、所述用户社交影响表征向量输入到因果预测模型中,利用所述反事实样本数据和相应的事实数据共同训练初始用户流失预测模型,得到建模社交影响因果效应的用户流失预测模型。

[0043] 进一步的,所述的用户流失预测装置,还包括:

[0044] 第一数据预处理单元,用于在将所述样本用户数据输入到初始用户流失预测模型中训练之前,预先对获取的原始样本用户数据中的离散变量进行独热编码,对所述原始样本用户数据中的连续变量进行归一化处理;

[0045] 第二数据预处理单元,用于将独热编码得到的编码结果与归一化处理得到的处理结果进行拼接后,利用全连接神经网络模型将初始样本用户数据特征映射到连续向量空间,得到连续特征向量表示的样本用户数据。

[0046] 进一步的,所述个人目标数据包含用户的基础属性数据和历史行为数据;所述关联关系数据是基于存在用户交互行为的应用程序对应的全量交互历史数据构建的关联关系网络得到的。

[0047] 第三方面,本发明还提供一种电子设备,包括:存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,所述处理器执行所述程序时实现如上述任意一项所述的用户流失预测方法的步骤。

[0048] 第四方面,本发明还提供一种处理器可读存储介质,所述处理器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执行时实现如上述任意一项所述的用户流失预测方法的步骤。

[0049] 采用本发明所述的用户流失预测方法,基于反事实推理建模架构引入用户间相互关系因素,能够有效适用于当前大量的基于社交网络或者存在用户间交互关系的应用程序,提高了用户流失预测的准确度和可解释性。

附图说明

[0050] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作一简单地介绍,显而易见地,下面描述中的附图是本发明的一些实施例,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取其他的附图。

[0051] 图1为本发明实施例提供的用户流失预测方法的流程示意图;

[0052] 图2为本发明实施例提供的因果假设的示意图

- [0053] 图3为本发明实施例提供的用户流失预测模型的架构示意图；
- [0054] 图4为本发明实施例提供的用户流失预测方法中的数据预处理示意图；
- [0055] 图5为本发明实施例提供的两路表征网络模型的架构示意图；
- [0056] 图6为本发明实施例提供的用户流失预测装置的结构示意图；
- [0057] 图7为本发明实施例提供的电子设备的实体结构示意图。

具体实施方式

[0058] 为使本发明实施例的目的、技术方案和优点更加清楚,下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获取的所有其他实施例,都属于本发明保护的范围。

[0059] 本发明提供一种在建模个人流失内在倾向的同时,有效地建模用户间关联关系影响的因果效应的用户流失预测模型,从而更精准地预测用户的流失情况。具体的,本发明是一种面向基于社交网络或者存在用户交互关系的应用平台,通过获取应用平台中用户间交互数据重构用户关联关系网络,并基于该用户关联关系网络数据、历史交互数据、基础属性数据、历史行为数据利用所述用户流失预测模型来预测用户流失的方法。本方法首先利用两个深度神经表征网络从获取的数据中分别提取出用户自身流失倾向表征向量和用户社交影响表征向量;之后,利用反事实数据预测子模型通过观测数据生成反事实数据;然后,将之前得到的两个表征向量输入因果预测模块,利用所述反事实数据及相应的事实数据共同训练初始用户流失预测模型,从而得到存在关于关联关系影响的因果效应的用户流失预测模型;最后,面向待预测用户,将关联关系数据、历史交互数据、个人目标数据输入训练完成的所述用户流失预测模型,得到对未来一段时间是否流失的预测结果。

[0060] 下面基于本发明所述的用户流失预测方法,对其实施例进行详细描述。如图1所示,其为本发明实施例提供的用户流失预测方法的流程示意图,具体实现过程包括以下步骤:

[0061] 步骤101:获取待预测用户的关联关系数据、历史交互数据及个人目标数据。其中,所述个人目标数据包含用户的基础属性数据(包括年龄、性别、注册时间、注册渠道等等)和历史行为数据等。所述用户交互信息可基于应用平台直接获取得到。需要说明的是关联关系数据的获取方法,对于存在用户交互行为(比如分享行为)的应用平台,可以应用全量交互历史数据构建的关联关系网络得到所述关联关系数据。比如只要两个用户在历史中交互过,即可认为两个用户之间存在关联关系,进而以此构建关联关系网络。

[0062] 在获取待预测用户的关联关系数据、历史交互数据及个人目标数据之前,需要预先对初始用户流失预测模型进行训练,得到满足需要的用户流失预测模型。

[0063] 在实际实施过程中,从应用平台获取的历史数据中提取到的原始个人目标数据包含:年龄、性别、注册时长、注册渠道等。历史交互数据包含每两个存在关联关系的用户之间的历史交互次数等。需要说明的是,以上数据为所有包含任意用户交互的应用平台均能得到的基础信息,应用时可在此基础信息上添加其他信息,在此不做具体限定。

[0064] 由于上述数据中既有离散变量(比如性别),也有连续变量(比如历史行为),为了让初始用户流失预测模型能够更好的学习各个变量与预测目标的关联关系以及变量之间

的交互影响,还需要预先对原始变量先进行编码或者归一化处理,并将其映射到连续向量空间。实现过程,如图4所示,对获取的原始样本用户数据中的离散变量进行独热编码,对所述原始样本用户数据中的连续变量进行归一化处理;将独热编码得到的编码结果与归一化处理得到的处理结果进行拼接后,利用全连接神经网络模型将初始样本用户数据特征映射到连续向量空间,得到连续特征向量表示的样本用户数据。比如首先对于离散变量,根据其取值范围进行独热编码,即将其包含的每一个取值表示为只有一个维度为1,其余维度为0的n维向量,其中n表示离散变量取值范围。对于连续离散变量,将其归一化到0到1的区间,从而减少特征尺度不同所带来的影响。将离散特征与连续特征拼接后,利用全连接神经网络模型将数据特征映射到连续向量空间,得到样本用户数据。

[0065] 获取预处理完成之后样本用户数据之后,可基于所述样本用户数据和标签用户流失数据对初始用户流失预测模型进行训练,得到需要的用户流失预测模型。具体的,将所述样本用户数据输入到初始用户流失预测模型中,利用残差深度交叉网络子模型从样本用户数据中提取出用户自身流失倾向表征向量,利用社交驱动的图神经网络子模型从样本用户数据中提取出用户社交影响表征向量;利用反事实数据预测子模型,基于相应的观测数据生成反事实样本数据;将所述用户自身流失倾向表征向量、所述用户社交影响表征向量输入到因果预测模型中,利用所述反事实样本数据和相应的事实数据共同训练初始用户流失预测模型,得到建模社交影响因果效应的用户流失预测模型。

[0066] 在具体实施过程中,首先可基于应用平台所保存的所有历史数据,获取过去一段时期中用户的关联关系数据、历史交互信息、基础属性数据和历史行为数据;之后,可以确定当前时期已经流失的用户,并以此作为模型训练的标签用户流失数据。将以上数据组成数据集,并按照8:2的比例划分为训练集和测试集进行训练,得到训练完成的用户流失预测模型。面向待预测用户,获取待预测用户当前阶段的关联关系数据、用户交互数据、基础属性数据和历史行为数据。将这些数据输入到训练完成的用户流失预测模型中,即能够得到用户流失预测模型对用户未来目标时间段是否会流失的预测结果。需要说明的是,随着时间推移,历史数据会不断积累,因此,可以每隔一段时间获取新的数据集,并对模型做进一步的训练,从而提升模型精度。

[0067] 具体而言,可以将用户流失预测模型表示为如下形式,已知信息包括:用户关联关系网络的结构 G (其中,用户被建模为节点,用户之间的关联关系被建模为边)、用户的个人目标数据(包括用户的基础属性数据与历史行为数据)特征矩阵 X_v ,用户间交互数据特征矩阵 X_e ,用户流失预测的任务目标是利用上述已知信息来学习一个映射函数 $F(\cdot)$ 来预测未来一段时间目标用户的流失情况 y :

$$[0068] \quad y = F(G, X_v, X_e)$$

[0069] 其中, $y = [y_1, y_2, \dots, y_N]^T$ 为用户流失预测模型对各个目标用户的预测结果, N 为目标用户总数。需要说明的是,在用户流失预测模型训练过程中需要对超参数进行搜索调优(超参数包括学习速率 β ,正则化系数 λ ,特征空间大小 d 等)。在对用户流失预测模型的训练过程中,本发明优先使用Adam优化器,该优化器是一种基于随机梯度下降的优化器,能够对用户流失预测模型中各层网络的参数进行更新。

[0070] 步骤102:将所述关联关系数据、所述历史交互数据及所述个人目标数据输入到用户流失预测模型中,得到所述用户流失预测模型输出的未来目标时间段对应的用户流失预

测结果。

[0071] 其中,所述用户流失预测模型是基于样本用户数据、所述样本用户数据对应的用户流失预测结果以及标签用户流失数据训练得到的。所述用户流失预测模型包含用于提取用户自身流失倾向表征向量的残差深度交叉网络子模型、用于提取用户社交影响表征向量的图神经网络子模型及用于生成反事实数据的反事实数据预测子模型。

[0072] 在本发明实施例中,所述的用户流失预测模型基于如图2所示的因果假设。具体而言,将用户流失的原因分为外源性的社交影响和内源性的用户倾向。所述的用户流失预测模型如图3所示。用户流失预测模型包含两个模块。第一个模块是一个两路的深度神经表征网络,其中包含残差深度交叉网络(Residual Deep and Cross Network,简称ResDCN)和基于社交驱动的图神经网络(Social Interaction Guided Graph Neural Network,简称SGAT)两个子模型,从获取的数据中先分别提取出包含用户自身流失倾向的表征向量 H_{UI} 和用户受到的社交影响的表征向量 H_{SI} 。第二个模块是反事实数据预测模块,即反事实数据预测子模型。该反事实数据预测子模型首先基于原始观测到的事实数据做反事实数据扩充,之后利用预设的三头预测的多任务预测器,让模型同时预测事实结果(Factual Outcome)、反事实结果(Counterfactual Outcome)和处理变量(Treatment),使得模型能够抓住社交影响对用户流失的因果效应。最终,用户流失预测模型输出的事实结果即为对用户流失的预测结果,即未来目标时间段对应的用户流失预测结果。模型中各模块详述如下:

[0073] 在数据预处理过程中,基于用户的个人目标数据中提取的个人特征 X_v 以及历史交互数据中提取的交互特征 X_e ,能够得到用户个人属性特征矩阵 $H_v = [h_{v_1}, h_{v_2}, \dots, h_{v_N}]^T \in \mathbb{R}^{N \times n_v}$ 和用户间交互特征矩阵 $H_e = [\dots, h_{e_{ij}}, \dots]^T \in \mathbb{R}^{L \times n_e}$,其中 N, L 分别为总用户数和总用户关系数, n_v, n_e 分别为用户个人属性向量维度和交互向量维度。

[0074] 为了从获取的数据中学习得到和用户自身倾向相关的信息,本发明提出了残差深度交叉网络子模型。具体而言,用户自身倾向可以从三类信息中获取,包括该用户是谁(基础属性数据)、该用户做过什么事(历史行为数据)以及用户有怎样的朋友圈(关联关系数据)。首先用一个图神经网络子模型提取出用户朋友圈的相关信息的嵌入向量 H_g ,将 H_g 与用户个人属性特征拼接后得到用户的特征 H 。其计算过程可以表示如下:

$$[0075] \quad H_g = \sigma(\widehat{A}H_vW_g),$$

$$[0076] \quad \widehat{A} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}} + I,$$

$$[0077] \quad H = H_g || H_v.$$

[0078] 其中, $(\cdot || \cdot)$ 是向量拼接操作, $\sigma(\cdot)$ 是激活函数,具体实现时可以采用ReLU等函数, D 为度矩阵, I 为单位阵, W_g 为模型参数。

[0079] 进一步,对于深度模型来说,特征之间的交叉项含有丰富的信息,能够很好的提升模型的精度。为此,本发明引入了深度交叉网络来进一步建模特征间的交叉关系,从而提升用户流失预测模型的表征能力。为了让用于流失预测模型更容易学习,本发明将深度交叉网络中的深度网络模块替换为了深度残差网络,所述深度残差网络,用于基于第1层交叉网络的隐节点向量和第1层深度网络的隐节点向量,得到第1+1层交叉网络的隐节点向量和第

1+1层深度网络的隐节点向量；并基于所述第1+1层交叉网络的隐节点向量和所述第1+1层深度网络的隐节点向量，得到所述用户自身流失倾向表征向量。所述深度残差网络的运算过程可以表示如下：

$$[0080] \quad \mathbf{H}_{\text{cross}}^{l+1} = \mathbf{H}_{\text{cross}}^l \odot (\mathbf{W}_{\text{cross}}^l \mathbf{H}_{\text{cross}}^l) + \mathbf{b}_{\text{cross}}^l$$

$$[0081] \quad \mathbf{H}_{\text{deep}}^{l+1} = \sigma(\mathbf{W}_{\text{deep}}^l \mathbf{H}_{\text{deep}}^l + \mathbf{b}_{\text{deep}}^l) + \mathbf{H}_{\text{deep}}^l$$

$$[0082] \quad \mathbf{H}_{\text{UI}} = \mathbf{H}_{\text{cross}}^{l+1} + \mathbf{H}_{\text{deep}}^{l+1}$$

[0083] 其中， \odot 代表按位点乘运算； $\mathbf{H}_{\text{cross}}^l$ 代表第1层交叉网络的隐节点向量， $\mathbf{H}_{\text{deep}}^l$ 代表第1层深度网络的隐节点向量； $\mathbf{W}_{\text{cross}}^l$ ， $\mathbf{W}_{\text{deep}}^l$ ， $\mathbf{b}_{\text{cross}}^l$ ， $\mathbf{b}_{\text{deep}}^l$ 代表模型参数。最终，残差深度交叉网络子模型的输出即为用户自身流失倾向的表征向量 \mathbf{H}_{UI} 。

[0084] 为了提取包含用户社交影响的信息，通常选择图神经网络(Graph Neural Networks, 简称GCN)。然而，现有技术中已有的GCN无法很好的刻画用户间的交互数据，而交互数据是社交影响重要的信息来源。为此，本发明设计了一个基于社交驱动的图神经网络子模型，其核心思想是每个节点根据其相邻节点的当前状态向量和历史交互数据，通过注意力机制更新自己的状态向量。

[0085] 所述图神经网络子模型中包含2个组件。首先是如图5中(b) SGAT部分的右侧所示的边学习组件。其中，所述边学习组件用于将输入的节点特征、边特征以及节点间交互特征进行拼接，并利用预设线性网络模型学习得到针对关联关系网络中每一条边的蒙版向量。具体的，此边学习组件的输入为节点特征 \mathbf{H}_v 和边特征 \mathbf{H}_e 。对于每一对相邻的节点 i 和 j ， $(i, j) \in E$ ， E 为网络中边的集合，此边学习组件首先将节点 i 的特征 \mathbf{h}_{v_i} 、节点 j 的特征 \mathbf{h}_{v_j} 、节点 i 与 j 间的交互特征 $\mathbf{h}_{e_{ij}}$ 拼接起来后，利用一个线性网络模型学习得到针对每一条边的蒙版向量 ϵ_{ij} 。该边学习组件的计算过程可以表示为：

$$[0086] \quad \epsilon_{ij} = \sigma(\mathbf{W}_{\epsilon}(\mathbf{h}_{v_i} || \mathbf{h}_{v_j} || \mathbf{h}_{e_{ij}}) + \mathbf{b}_{\epsilon})$$

[0087] 其中， $(\cdot || \cdot)$ 是向量拼接操作； $\sigma(\cdot)$ 是激活函数；具体实现时可以采用ReLU等函数； \mathbf{W}_{ϵ} ， \mathbf{b}_{ϵ} 为模型参数。需要注意的是 $\epsilon_{ij} \neq \epsilon_{ji}$ 。

[0088] 第二个组件是基于社交驱动的图卷积组件，该图卷积组件用于针对所述关联关系网络中的目标节点，基于所述边学习组件得到的对应蒙版向量以及相邻节点的状态向量，通过注意力机制更新所述目标节点的状态向量；将所述目标节点的状态向量作为所述用户社交影响表征向量。其核心创新是在传统GCN卷积的邻居节点聚合的步骤中加入了一个基于边的控制器。使得每个节点基于边学习得到蒙版向量 ϵ_{ij} 以及其邻居状态向量 \mathbf{h}_{v_j} ，并通过注意力机制更新节点自身的状态信息。所述图卷积组件的计算过程可以表示为：

$$[0089] \quad \mathbf{h}_{v_i}^{l+1} = \alpha_{ii}^l \mathbf{W}_s^l \mathbf{h}_{v_i}^l + \sum_{j \in \mathcal{N}(i)} \alpha_{ij}^l \mathbf{W}_s^l (\epsilon_{ij}^l \odot \mathbf{h}_{v_j}^l)$$

$$[0090] \quad \alpha_{ij}^l = \text{softmax} \left(\sigma \left(\mathbf{a}^T \left[\mathbf{W}_s^l \mathbf{h}_{v_i}^l \parallel \mathbf{W}_s^l \left(\epsilon_{ij}^l \odot \mathbf{h}_{v_j}^l \right) \right] \right) \right)$$

[0091] 其中 $\mathbf{h}_{v_i}^{l+1}$ 为节点 i 更新后的状态向量; $\sigma(\cdot)$ 是激活函数; $\mathcal{N}(i)$ 为返回节点 i 的邻居节点集合的函数; \odot 代表按位点乘运算; $\mathbf{a}^T, \mathbf{W}_s^l$ 为模型参数。

[0092] 需要说明的是,图5中,本发明只以一层SGAT网络作为示例,在实际应用中,可以堆叠多层SGAT网络以提升用户流失预测模型的表达能力。SGAT网络的输出即为用户流失预测模型学习到的社交影响的表征向量 \mathbf{H}_{SI} 。

[0093] 为了更为清晰地介绍本发明的反事实数据预测框架,本发明首先介绍一些因果推断的术语。在本发明实施例中,采用Rubin所提出的被广泛使用的因果推断框架——潜在结果框架。此框架中,谈论因果性都是对一个个体或者一个群体来说,某种处理(Treatment)对个体或者群体的影响。比如处理可以指接受药物等干预。基于用户流失预测的场景,本发明定义处理为用户是否有存在关联关系的用户流失。当个体接受处理时,即有朋友流失时,处理变量 $t=1$;反之 $t=0$ 。本发明记处理变量为0时观测到的结果(Outcome)为 \mathbf{y}_{t_0} ,处理变量为1时观测到的结果为 \mathbf{y}_{t_1} 。在用户流失预测场景,结果即指目标用户是否流失。由于对于一个个体,其要么有存在关联关系的用户流失,要么没有,不可能处于一个中间态。因此,对于同一个个体,本发明只能观测到 \mathbf{y}_{t_0} 和 \mathbf{y}_{t_1} 中的一个结果,不可能同时观测到二者。本发明记观测到的个体的事实数据结果(Factual Outcome)为变量 y_f ,未观测到的反事实数据结果(Counterfactual Outcome)为变量 y_{cf} 。以上数据变量的关系可以由下式表示:

$$[0094] \quad y_f = t \times y_{t_1} + (1 - t) \times y_{t_0},$$

$$[0095] \quad y_{cf} = t \times y_{t_0} + (1 - t) \times y_{t_1}.$$

[0096] 本发明所提出的反事实数据预测子模型包含两个模块,分别是反事实数据扩增(Counterfactual Data Augmentation)模块和反事实数据预测(Counterfactual Prediction)模块。所述反事实数据扩增模块,用于基于预设的因果关系模型以及反事实数据扩增推论规则,得到反事实数据;所述反事实数据预测模块,用于通过预设的三头预测的多任务预测器对所述反事实数据、相应的事实数据以及处理变量进行预测,基于因果正则项建模社交影响因果效应,以得到对应的用户流失预测结果。本发明对各模块详述如下。

[0097] 针对反事实数据扩增模块:实践中深度模型捕捉因果效应的一个关键困难就是无法观测到反事实数据。为了解决这个困难,本发明基于图2中所示的因果图模型导出一个推论,并基于推论提出了基于反事实数据扩增推论规则的方法。具体来说,由于去掉任何一个原因只会降低结果的概率,因此本发明可以得到:目标用户在有关联关系的用户(比如与目标用户之间存在朋友关系的用户)流失的时候,该目标用户流失的概率将大于没有关联关系的用户流失的时候。此推论对应的表达式可以表示为:

$$[0098] \quad P(y_{t_1}) \geq P(y_{t_0})$$

[0099] 根据这一推论,本发明可以进一步得到以下两个推论:

[0100] (1) 如果一个用户在没有关联关系的用户流失的情况下流失了 ($t=0, y_f=1$),那么他在有关联关系的用户流失的情况下,同样会流失 ($t=1, y_{cf}=1$)。(2) 如果一个用户在有

关联关系的用户流失的情况下也没有流失 ($t=1, y_f=0$), 那么他在没有关联关系的用户流失的情况下同样不会流失 ($t=0, y_{cf}=0$)。由此本发明可以得到反事实数据集 \mathcal{O}_{cf} 。此过程可以表示如下:

$$[0101] \quad \mathcal{O}_{cf1} = \{y_{cf}^i = y_{t_1}^i = 1 \mid y_f^i = y_{t_0}^i = 1, i \in \mathcal{O}_f\},$$

$$[0102] \quad \mathcal{O}_{cf2} = \{y_{cf}^i = y_{t_0}^i = 0 \mid y_f^i = y_{t_1}^i = 0, i \in \mathcal{O}_f\},$$

$$[0103] \quad \mathcal{O}_{cf} = \mathcal{O}_{cf1} \cup \mathcal{O}_{cf2},$$

[0104] 针对反事实数据预测模块: 在得到反事实数据后, 本发明可基于预设的三头的多任务预测器来同时预测事实数据、反事实数据和处理变量。具体来说, 如图3所示, 本发明根据之前得到的用户自身倾向表征向量 H_{UI} 和用户所受社交影响的表征向量 H_{SI} , 通过注意力机制分别预测用户受到处理和没有受到处理时是否流失, 即预测 \mathbf{y}_{t_0} 和 \mathbf{y}_{t_1} 。进一步的, 可以通过 \mathbf{y}_{t_0} 、 \mathbf{y}_{t_1} 与 y_f 、 y_{cf} 的关系得到事实数据和反事实数据的预测。预测网络可以表示如下:

$$[0105] \quad \hat{\mathbf{y}} = \text{sigmoid} \left(\mathbf{p}^T \sigma \left(\mathbf{W}_{fc} \text{attention} (H_{SI} \parallel H_{UI}) + \mathbf{b}_{fc} \right) \right)$$

[0106] 其中, $\hat{\mathbf{y}}$ 为预测变量 (可以对应 \mathbf{y}_{t_0} 或 \mathbf{y}_{t_1}); $\mathbf{p}^T, \mathbf{W}_{fc}, \mathbf{b}_{fc}$ 为模型参数。进一步, 本发明构建的事实预测损失函数、反事实数据预测损失函数如下:

$$[0107] \quad \mathcal{L}_f(y_f, \hat{y}_f) = \frac{1}{N} \sum_{i \in \mathcal{O}_f} y_f^i \log(\hat{y}_f^i) + (1 - y_f^i) \log(1 - \hat{y}_f^i)$$

$$[0108] \quad \mathcal{L}_{cf}(y_{cf}, \hat{y}_{cf}) = \frac{1}{M} \sum_{i \in \mathcal{O}_{cf}} y_{cf}^i \log(\hat{y}_{cf}^i) + (1 - y_{cf}^i) \log(1 - \hat{y}_{cf}^i)$$

[0109] 其中, N 为观测数据集用户数, M 为反事实数据用户数。

[0110] 基于倾向得分的充分性理论 (The sufficiency of the propensity score), 在预测事实数据和反事实数据的同时, 预测用户所受到的处理能够让用户流失预测模型建模社交影响的因果效应。因此, 本发明基于 H_{SI} , 利用一个全连接神经网络预测用户受到的处理, 即用户是否有存在关联关系的用户流失。处理变量损失函数可以表示如下:

$$[0111] \quad \mathcal{L}_t(t, \hat{t}) = \frac{1}{N + M} \sum_{i \in \mathcal{O}_f \cup \mathcal{O}_{cf}} t^i \log(\hat{t}^i) + (1 - t^i) \log(1 - \hat{t}^i)$$

[0112] 进一步, 如图5所示, 本发明引入因果正则项以帮助模型更好的建模因果效应。具体来说, 对于同一个目标用户本发明强制让用户流失预测模型预测的有处理时的流失概率 \mathbf{y}_{t_1} 大于或等于其没有受到处理时的流失概率 \mathbf{y}_{t_0} 。由此构建损失函数如下:

$$[0113] \quad \mathcal{L}_c(\hat{y}_{t_0}, \hat{y}_{t_1}) = \frac{1}{M} \sum_{i \in \mathcal{O}_f \cup \mathcal{O}_{cf}} \max(0, \hat{y}_{t_0}^i - \hat{y}_{t_1}^i)$$

[0114] 其中, $\max(\cdot, \cdot)$ 为取最大值函数。

[0115] 另外, 考虑到实际实施过程中, 用户流失是一个存在较大数据不均衡的问题, 即流

失用户往往远少于非流失用户,导致模型训练的正样本远小于负样本,进而模型在训练中容易忽视正样本,训练效果差。为了使得训练过程更为重视正样本,本发明为其加上权重 α_d ,同时为了让模型在各个任务中取得平衡,取得对用户流失最好的预测效果,本发明对各个预测任务都加入一个可调的超参数 $\alpha_{cf}, \alpha_t, \alpha_c$ 作为权重。由此,本发明最终的损失函数可以表示如下:

$$[0116] \quad \mathcal{L} = (1 + y_f \times \alpha_d) \times (\mathcal{L}_f + \alpha_{cf} \mathcal{L}_{cf}) + \alpha_t \mathcal{L}_t + \alpha_c \mathcal{L}_c。$$

[0117] 下面以基于2020年三、四季度用户的历史交互数据、用户购买数据以及个人目标数据,预测A应用平台用户在2021年一季度期间是否会流失为例对本发明提供的用户流失预测方法进行说明:第一步,首先需要根据A应用平台的历史数据,获取得到用户的基础属性数据和历史交互数据。其中,基础属性数据具体包括:年龄、性别、注册渠道、注册时长、用户在每个月中购买商品的件数和金额等。需要注意的是,若存在部分用户没有填写年龄和性别信息,可将此类用户按照第三类别处理(比如性别的取值包括三类:男性、女性、未填写)。历史交互信息包括用户每个月中的交互次数等。第二步,训练模型。基于2020年第三季度的数据和第四季度观测到用户是否流失的结果构建训练数据集。将所有用户数据随机按照8:2分为训练集和测试集,将训练集输入到初始用户流失预测模型中训练;通过随机搜索的方式调整超参数,并利用测试集筛选出预测准确率最高、泛化能力最强的超参数对应的模型作为最终的用户流失预测模型。第三步,将2020年四季度用户的基础属性数据和用户的历史交互数据输入训练好的用户流失预测模型,得到2021年一季度用户的流失预测结果。

[0118] 采用本发明实施例所述的用户流失预测方法,基于反事实推理建模架构引入用户间相互关系因素,能够有效适用于当前大量的基于社交网络或者存在用户间交互关系的应用程序,提高了用户流失预测的准确度和可解释性。

[0119] 与上述提供的一种用户流失预测方法相对应,本发明还提供一种用户流失预测装置。由于该装置的实施例相似于上述方法实施例,所以描述得比较简单,相关之处请参见上述方法实施例部分的说明即可,下面描述的用户流失预测装置的实施例仅是示意性的。请参考图6所示,其为本发明实施例提供的一种用户流失预测装置的结构示意图。

[0120] 本发明所述的用户流失预测装置具体包括如下部分:

[0121] 用户数据获取单元601,用于获取待预测用户的关联关系数据、历史交互数据及个人目标数据;

[0122] 用户流失预测单元602,用于将所述关联关系数据、所述历史交互数据及所述个人目标数据输入到用户流失预测模型中,得到所述用户流失预测模型输出的未来目标时间段对应的用户流失预测结果。

[0123] 其中,所述用户流失预测模型是基于样本用户数据、所述样本用户数据对应的用户流失预测结果以及标签用户流失数据训练得到的;

[0124] 所述用户流失预测模型包含用于提取用户自身流失倾向表征向量的残差深度交叉网络子模型、用于提取用户社交影响表征向量的图神经网络子模型及用于生成反事实数据的反事实数据预测子模型。

[0125] 采用本发明实施例所述的用户流失预测装置,基于反事实推理建模架构引入用户

间相互关系因素,能够有效适用于当前大量的基于社交网络或者存在用户间交互关系的应用程序,提高了用户流失预测的准确度和可解释性。

[0126] 与上述提供的用户流失预测方法相对应,本发明还提供一种电子设备。由于该电子设备的实施例类似于上述方法实施例,所以描述得比较简单,相关之处请参见上述方法实施例部分的说明即可,下面描述的电子设备仅是示意性的。如图7所示,其为本发明实施例公开的一种电子设备的实体结构示意图。该电子设备可以包括:处理器(processor)701、存储器(memory)702和通信总线703,其中,处理器701,存储器702通过通信总线703完成相互间的通信,通过通信接口704与外部进行通信。处理器701可以调用存储器702中的逻辑指令,以执行用户流失预测方法,该方法包括:获取待预测用户的关联关系数据、历史交互数据及个人目标数据;将所述关联关系数据、所述历史交互数据及所述个人目标数据输入到用户流失预测模型中,得到所述用户流失预测模型输出的未来目标时间段对应的用户流失预测结果;其中,所述用户流失预测模型是基于样本用户数据、所述样本用户数据对应的用户流失预测结果以及标签用户流失数据训练得到的;所述用户流失预测模型包含用于提取用户自身流失倾向表征向量的残差深度交叉网络子模型、用于提取用户社交影响表征向量的图神经网络子模型及用于生成反事实数据的反事实数据预测子模型。

[0127] 此外,上述的存储器702中的逻辑指令可以通过软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个计算机可读取存储介质中。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,或者网络设备等)执行本发明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:存储芯片、U盘、移动硬盘、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、磁碟或者光盘等各种可以存储程序代码的介质。

[0128] 另一方面,本发明实施例还提供一种计算机程序产品,所述计算机程序产品包括存储在处理器可读存储介质上的计算机程序,所述计算机程序包括程序指令,当所述程序指令被计算机执行时,计算机能够执行上述各方法实施例所提供的用户流失预测方法,该方法包括:获取待预测用户的关联关系数据、历史交互数据及个人目标数据;将所述关联关系数据、所述历史交互数据及所述个人目标数据输入到用户流失预测模型中,得到所述用户流失预测模型输出的未来目标时间段对应的用户流失预测结果;其中,所述用户流失预测模型是基于样本用户数据、所述样本用户数据对应的用户流失预测结果以及标签用户流失数据训练得到的;所述用户流失预测模型包含用于提取用户自身流失倾向表征向量的残差深度交叉网络子模型、用于提取用户社交影响表征向量的图神经网络子模型及用于生成反事实数据的反事实数据预测子模型。

[0129] 又一方面,本发明实施例还提供一种处理器可读存储介质,所述处理器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执行时实现以执行上述各实施例提供的用户流失预测方法,该方法包括:获取待预测用户的关联关系数据、历史交互数据及个人目标数据;将所述关联关系数据、所述历史交互数据及所述个人目标数据输入到用户流失预测模型中,得到所述用户流失预测模型输出的未来目标时间段对应的用户流失预测结果;其中,所述用户流失预测模型是基于样本用户数据、所述样本用户数据对应的用户流失预

测结果以及标签用户流失数据训练得到的;所述用户流失预测模型包含用于提取用户自身流失倾向表征向量的残差深度交叉网络子模型、用于提取用户社交影响表征向量的图神经网络子模型及用于生成反事实数据的反事实数据预测子模型。

[0130] 所述处理器可读存储介质可以是处理器能够存取的任何可用介质或数据存储设备,包括但不限于磁性存储器(例如软盘、硬盘、磁带、磁光盘(MO)等)、光学存储器(例如CD、DVD、BD、HVD等)、以及半导体存储器(例如ROM、EPROM、EEPROM、非易失性存储器(NAND FLASH)、固态硬盘(SSD))等。

[0131] 以上所描述的装置实施例仅仅是示意性的,其中所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。本领域普通技术人员在不付出创造性的劳动的情况下,即可以理解并实施。

[0132] 通过以上的实施方式的描述,本领域的技术人员可以清楚地了解到各实施方式可借助软件加必需的通用硬件平台的方式来实现,当然也可以通过硬件。基于这样的理解,上述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ROM/RAM、磁碟、光盘等,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备等)执行各个实施例或者实施例的某些部分所述的方法。

[0133] 最后应说明的是:以上实施例仅用以说明本发明的技术方案,而非对其限制;尽管参照前述实施例对本发明进行了详细的说明,本领域的普通技术人员应当理解:其依然可以对前述各实施例所记载的技术方案进行修改,或者对其中部分技术特征进行等同替换;而这些修改或者替换,并不使相应技术方案的本质脱离本发明各实施例技术方案的精神和范围。

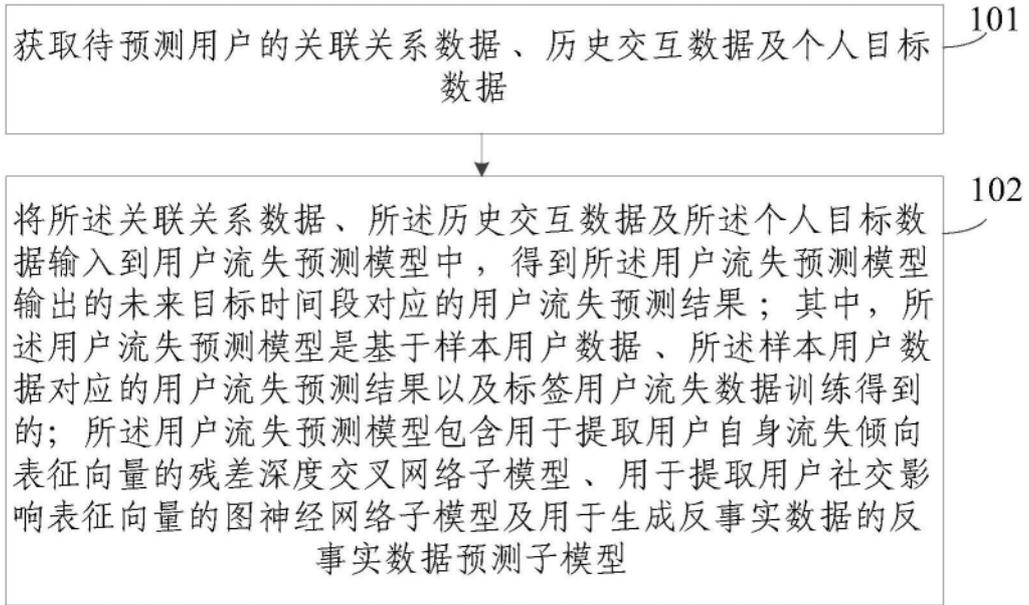


图1

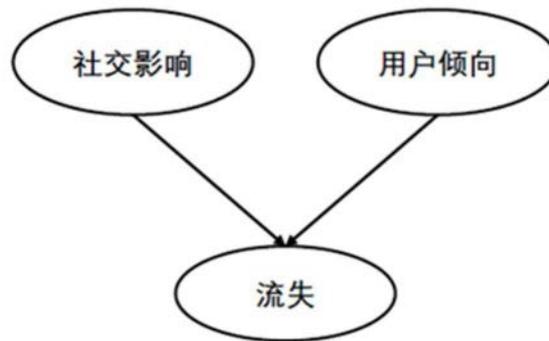


图2

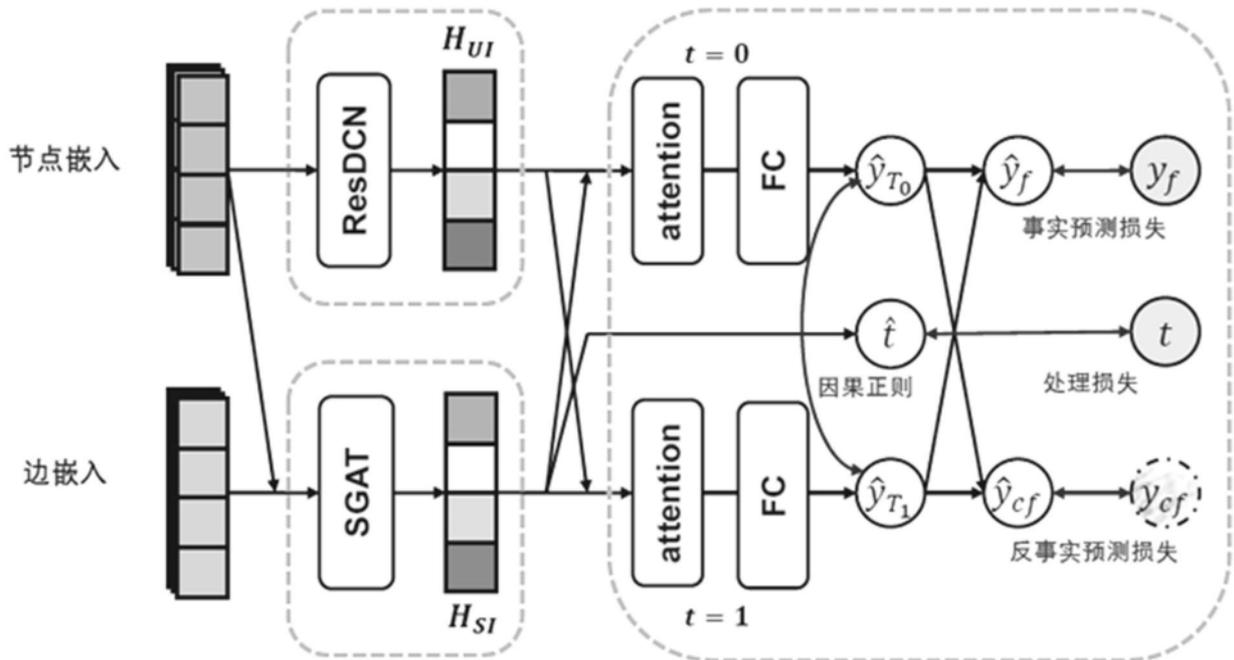


图3

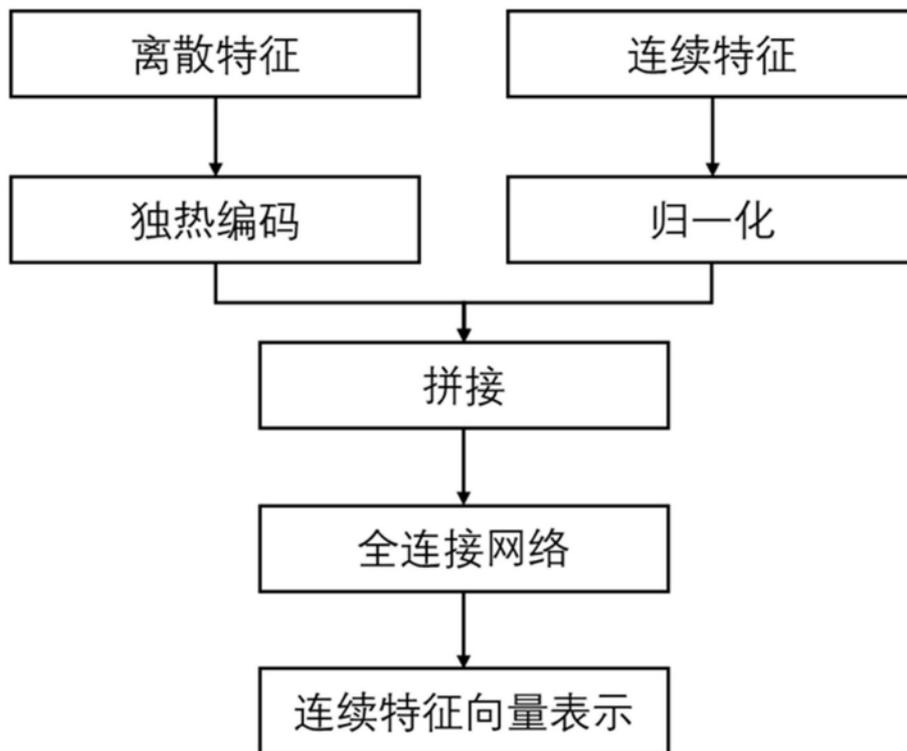


图4

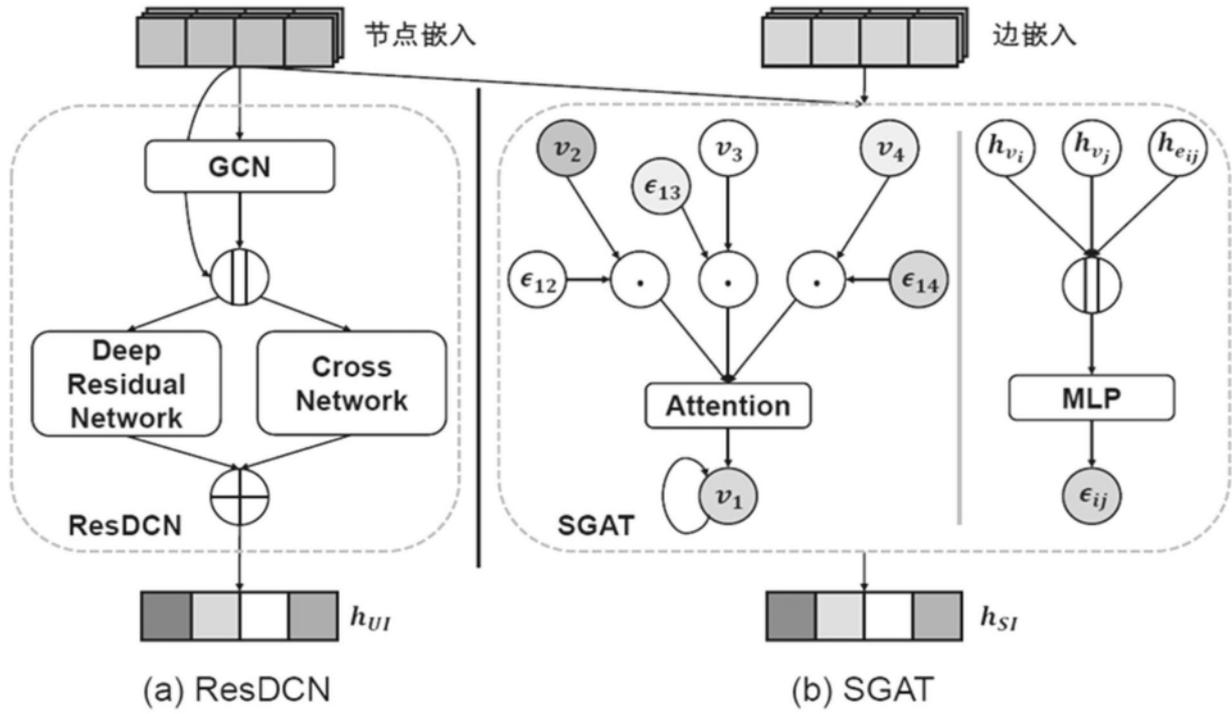


图5

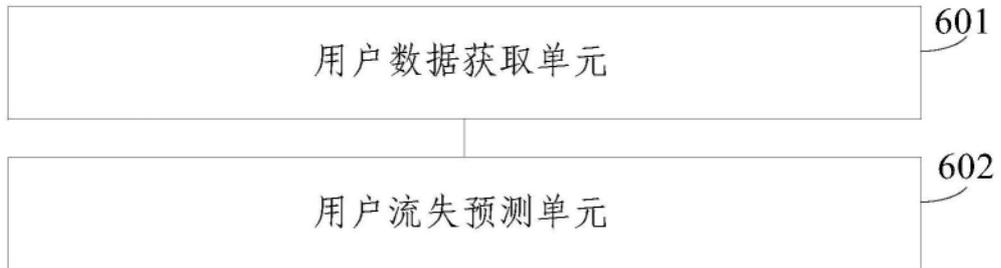


图6

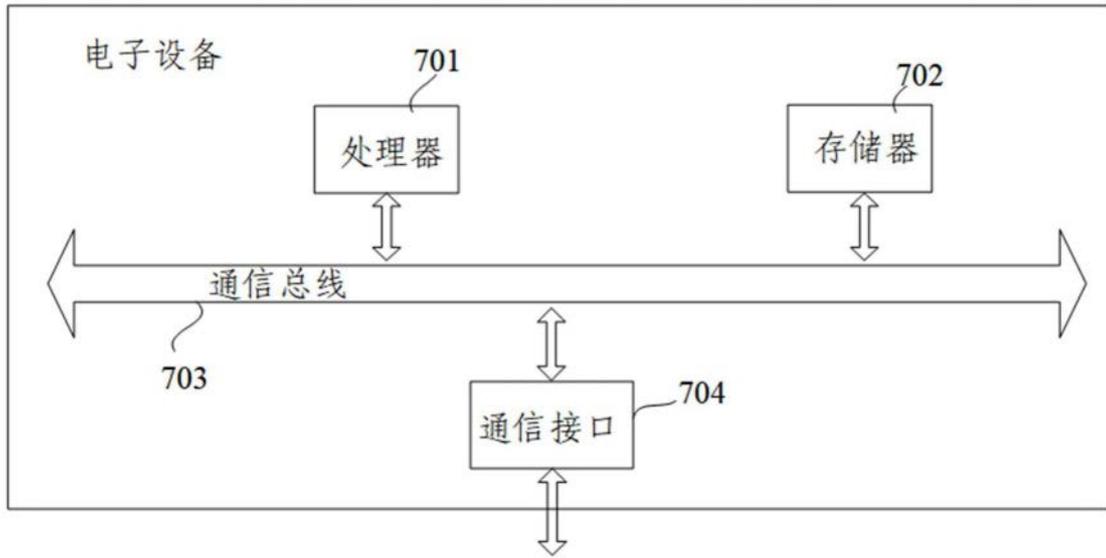


图7