

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020 年 7 月 9 日 (09.07.2020)



(10) 国际公布号
WO 2020/140639 A1

(51) 国际专利分类号:
G06F 16/901 (2019.01)

(21) 国际申请号: PCT/CN2019/119480

(22) 国际申请日: 2019 年 11 月 19 日 (19.11.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910002951.5 2019 年 1 月 2 日 (02.01.2019) CN

(71) 申请人: 平安科技(深圳)有限公司(PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN];
中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(72) 发明人: 徐凯(XU, Kai); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(74) 代理人: 深圳市明日今典知识产权代理事务所(普通合伙)(SHENZHEN MINGRIJINDIAN INTELLECTUAL PROPERTY AGENCY FIRM (GENERAL)); 中国广东省深圳市南山区南头街道安乐社区关口二路 15 号智恒产业园 30 栋 4 层 406 室, Guangdong 518000 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX,

(54) **Title:** MACHINE LEARNING-BASED REPORT GENERATING METHOD, APPARATUS, AND COMPUTER DEVICE

(54) 发明名称: 基于机器学习的报表生成方法、装置和计算机设备

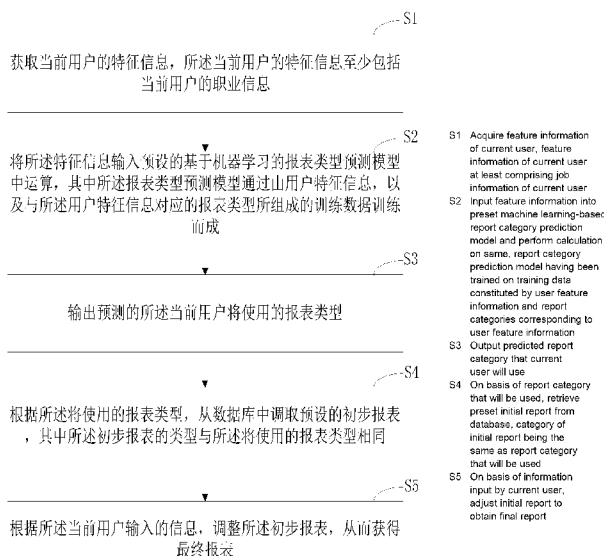


图 1

(57) **Abstract:** A machine learning-based report generating method and an apparatus, a computer device, and a storage medium, the method comprising: acquiring feature information of a current user; inputting the feature information into a preset machine learning-based report category prediction model and performing calculation on same, the report category prediction model having been trained on training data constituted by user feature information and report categories corresponding to the user feature information; outputting a predicted report category the current user will use; on the basis of the report category that will be used, retrieving a preset initial report from a database; and on the basis of information input by the current user, adjusting the initial report to obtain a final report. The report that the user requires is thereby correctly generated, reducing time and effort spent by the user in adjusting the report.

MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL,
PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,
SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国 (除另有指明, 要求每一种可提供的地区
保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ,
NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM,
AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG,
CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,
IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,
RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告 (条约第21条(3))。

(57) 摘要: 一种基于机器学习的报表生成方法、装置、计算机设备和存储介质, 所述方法包括: 获取当前用户的特征信息; 将所述特征信息输入预设的基于机器学习的报表类型预测模型中运算, 其中所述报表类型预测模型通过由用户特征信息, 以及与所述用户特征信息对应的报表类型所组成的训练数据训练而成; 输出预测的所述当前用户将使用的报表类型; 根据所述将使用的报表类型, 从数据库中调取预设的初步报表; 根据所述当前用户输入的信息, 调整所述初步报表, 从而获得最终报表。从而准确生成用户需要的报表, 减少用户调整报表的时间与精力。

基于机器学习的报表生成方法、装置和计算机设备

[0001] 本申请要求于2019年1月2日提交中国专利局、申请号为2019100029515，发明名称为“基于机器学习的报表生成方法、装置和计算机设备”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

[0002] 本申请涉及到计算机领域，特别是涉及到一种基于机器学习的报表生成方法、装置、计算机设备和存储介质。

背景技术

[0003] 在不同行业中，经常需要不同种类的报表，即使在同一个行业中，也可能需要不同种类的报表。现有技术不能有效地预测用户具体需要什么样的报表，因此经常需要用户手动调整报表类型、报表参数，从而需要繁琐的人工操作，降低用户的工作效率，用户体验必然极差。另外，现有技术的用户需要人工调整报表，以获得最终想要的报表，即没有报表模板、或者没有合适的报表模板，进而在需要新的报表的时候，不得不采用手动调整的方式，大幅度地调整报表。因此，现有技术的报表生成方案，不能避免用户花费额外调整报表的时间与精力。

发明概述

技术问题

[0004] 本申请的主要目的为提供一种基于机器学习的报表生成方法、装置、计算机设备和存储介质，旨在利用模板准确生成用户需要的报表，减少用户调整报表的时间与精力。

问题的解决方案

技术解决方案

[0005] 为了实现上述目的，本申请提出一种基于机器学习的报表生成方法，包括以下步骤：

[0006] 获取当前用户的特征信息，所述当前用户的特征信息至少包括当前用户的职业

信息;

- [0007] 将所述特征信息输入预设的基于机器学习的报表类型预测模型中运算, 其中所述报表类型预测模型通过由用户特征信息, 以及与所述用户特征信息对应的报表类型所组成的训练数据训练而成;
- [0008] 输出预测的所述当前用户将使用的报表类型;
- [0009] 根据所述将使用的报表类型, 从数据库中调取预设的初步报表, 其中所述初步报表的类型与所述将使用的报表类型相同;
- [0010] 根据所述当前用户输入的信息, 调整所述初步报表, 从而获得最终报表。
- [0011] 进一步地, 所述根据所述将使用的报表类型, 从数据库中调取预设的初步报表, 其中所述初步报表的类型与所述将使用的报表类型相同的步骤, 包括:
- [0012] 根据所述将使用的报表类型, 从预设的数据库中调取预设的多个图表模板与多个文字部分模板;
- [0013] 将所述当前用户选择的图表模板与文字部分模板组合成为所述初步报表;
- [0014] 调取所述初步报表。
- [0015] 进一步地, 所述根据所述当前用户输入的信息, 调整所述初步报表, 从而获得最终报表的步骤, 包括:
- [0016] 根据所述当前用户输入的图表调整信息、图表数据内容信息和文字部分调整信息, 调整所述初步报表中的图表和文字部分;
- [0017] 将所述当前用户输入的文字内容填入所述初步报表的文字部分, 获得最终报表。
- [0018] 进一步地, 所述报表类型预测模型的获取方法, 包括:
- [0019] 获取包括指定量的样本数据的训练集, 其中, 所述样本数据包括用户特征信息, 以及与所述用户特征信息对应的报表类型;
- [0020] 将所述训练集的样本数据输入到神经网络模型中进行训练, 其中, 训练的过程中采用随机梯度下降法, 利用反向传导法则更新所述神经网络模型各层的参数, 得到初步训练模型;
- [0021] 将所述初步训练模型记为所述报表类型预测模型。
- [0022] 进一步地, 所述将所述初步训练模型记为所述报表类型预测模型的步骤, 包括

:

- [0023] 获取包括指定量的样本数据的验证集，其中，所述验证集的样本数据包括用户特征信息，以及与所述用户特征信息对应的报表类型；
- [0024] 利用所述验证集的样本数据验证所述初步训练模型；
- [0025] 如果验证通过，则将所述初步训练模型记为所述报表类型预测模型。
- [0026] 进一步地，所述报表类型预测模型的获取方法，包括：
- [0027] 获取指定量的样本数据，其中，所述样本数据包括用户特征信息，以及与所述用户特征信息对应的报表类型；
- [0028] 将所述训练集的样本数据输入到CHAID决策树模型中进行训练，得到初步CHAID决策树；
- [0029] 将所述初步CHAID决策树记为所述报表类型预测模型。
- [0030] 进一步地，所述将所述训练集的样本数据输入到CHAID决策树模型中进行训练，得到初步CHAID决策树的步骤，包括：
- [0031] 设置所述CHAID决策树模型的建模标准参数，所述建模标准参数包括决策树的最大层数、母节点的可再分的显著水平、母节点包含的最小样本数和子节点包含的最小样本数；
- [0032] 将训练集的样本数据输入到采用卡方自动交互检测法建立的所述CHAID决策树模型中进行训练，得到初步CHAID决策树。
- [0033] 本申请提供一种基于机器学习的报表生成装置，包括：
- [0034] 特征信息获取单元，用于获取当前用户的特征信息，所述当前用户的特征信息至少包括当前用户的职业信息；
- [0035] 报表类型预测模型运算单元，用于将所述特征信息输入预设的基于机器学习的报表类型预测模型中运算，其中所述报表类型预测模型通过由用户特征信息，以及与所述用户特征信息对应的报表类型所组成的训练数据训练而成；
- [0036] 报表类型预测单元，用于输出预测的所述当前用户将使用的报表类型；
- [0037] 初步报表调取单元，用于根据所述将使用的报表类型，从数据库中调取预设的初步报表，其中所述初步报表的类型与所述将使用的报表类型相同；
- [0038] 最终报表获得单元，用于根据所述当前用户输入的信息，调整所述初步报表，

从而获得最终报表。

[0039] 本申请提供一种计算机设备，包括存储器和处理器，所述存储器存储有计算机可读指令，所述处理器执行所述计算机可读指令时实现上述任一项所述方法的步骤。

[0040] 本申请提供一种计算机可读存储介质，其上存储有计算机可读指令，所述计算机可读指令被处理器执行时实现上述任一项所述的方法的步骤。

发明的有益效果

有益效果

[0041] 本申请的基于机器学习的报表生成方法、装置、计算机设备和存储介质，通过基于机器学习的报表类型预测模型，以预测当前用户将使用的报表类型，再根据所述报表类型，从数据库中调取预设的初步报表，对初步报表进行调整后以得到最终报表，从而避免用户的繁琐操作，提升用户体验，提高报表完成效率。

对附图的简要说明

附图说明

[0042] 图1 为本申请一实施例的基于机器学习的报表生成方法的流程示意图；

[0043] 图2 为本申请一实施例的基于机器学习的报表生成装置的结构示意框图；

[0044] 图3 为本申请一实施例的计算机设备的结构示意框图。

[0045] 本申请目的的实现、功能特点及优点将结合实施例，参照附图做进一步说明。

实施该发明的最佳实施例

本发明的最佳实施方式

[0046] 为了使本申请的目的、技术方案及优点更加清楚明白，以下结合附图及实施例，对本申请进行进一步详细说明。应当理解，此处描述的具体实施例仅仅用以解释本申请，并不用于限定本申请。

[0047] 参照图1，本申请实施例提供一种基于机器学习的报表生成方法，包括以下步骤：

[0048] S1、获取当前用户的特征信息，所述当前用户的特征信息至少包括当前用户的

职业信息；

- [0049] S2、将所述特征信息输入预设的基于机器学习的报表类型预测模型中运算，其中所述报表类型预测模型通过由用户特征信息，以及与所述用户特征信息对应的报表类型所组成的训练数据训练而成；
- [0050] S3、输出预测的所述当前用户将使用的报表类型；
- [0051] S4、根据所述将使用的报表类型，从数据库中调取预设的初步报表，其中所述初步报表的类型与所述将使用的报表类型相同；
- [0052] S5、根据所述当前用户输入的信息，调整所述初步报表，从而获得最终报表。
- [0053] 如上述步骤S1所述，获取当前用户的特征信息，所述当前用户的特征信息至少包括当前用户的职业信息。其中，所述当前用户的特征信息是指，能够体现当前用户的特性的信息，例如当前用户的职业、当前用户近期（指定时间内）使用的报表类型、当前用户的年龄、当前用户的性别。其中，当前用户的职业信息与其可能使用的报表类型有较大关系，所以所述当前用户的特征信息至少包括当前用户的职业信息，以提高报表类型预测的准确性。例如，金融行业中的股票评论师，有较高机率使用股票报表。获取当前用户的特征信息的过程包括：从所述当前用户的注册帐户信息中提取所述当前用户的特征信息。
- [0054] 如上述步骤S2所述，将所述特征信息输入预设的基于机器学习的报表类型预测模型中运算，其中所述报表类型预测模型通过用户特征信息，及与所述用户特征信息对应的报表类型训练而成。基于机器学习的报表类型预测模型，通过不断的自我学习，提高预测的准确性，避免预测出错误的报表类型。所述报表类型预测模型可以基于任意机器学习的模型生成，例如神经网络模型、分类树模型，再经由训练数据训练而成。其中，所述报表类型预测模型通过由用户特征信息，以及与所述用户特征信息对应的报表类型所组成的训练数据训练而成。所述用户特征信息可以包括任意能够体现用户的特性的信息，例如用户的职业、用户近期（指定时间内）使用的报表类型、用户的年龄、用户的性别。其中，所述用户指使用报表的用户。其中，所述报表类型预测模型用于根据用户特征信息，预测出用户将要使用的报表类型。
- [0055] 如上述步骤S3所述，输出预测的所述当前用户将使用的报表类型。所述报表类

型可以为任意分类方法分成的类型，例如以包括的图表种类来分为：饼状图报表（报表中包括饼状图）、曲线图报表、直方图报表；根据用途来分为：财务报表、金融报表、统计分析报表。通过所述报表类型预测模型，即可输出预测的所述当前用户将使用的报表类型。

[0056] 如上述步骤S4所述，根据所述将使用的报表类型，从数据库中调取预设的初步报表，其中所述初步报表的类型与所述将使用的报表类型相同。其中预设的数据库中预存有不同报表类型的初步报表，同一报表类型的初步报表可以有多种，以供用户选择。例如饼状图报表初步报表可以包括饼状图位于报表顶部、位于报表中间、位于报表末尾等三种报表；金融初步报表可以包括具有K线图的报表、具有直方图的报表、具有曲线图的报表等。这些初步报表可供用户直接选择，从而省去了用户琐碎地从零开始建立报表的过程，提高了报表形成速度。其中，所述预设的初步报表可以是已经建好的完整的报表，也可以是由当前用户从多个备选的图表模板与文字部分模板中选出需要的确切的图表模板与文字部分模板所组合而成的报表。

[0057] 如上述步骤S5所述，根据所述当前用户输入的信息，调整所述初步报表，从而获得最终报表。由前述，提供了初步报表，再根据所述当前用户输入的信息，即可得到最终报表。其中，所述当前用户输入的信息，包括对初步报表的参数进行调整的信息、对初步报表进行描述的文字内容信息中的至少一种。其中，所述最终报表至少包括图表部分，进一步地，还可以包括文字部分。

[0058] 在一个实施方式中，所述根据所述将使用的报表类型，从数据库中调取预设的初步报表，其中所述初步报表的类型与所述将使用的报表类型相同的步骤S4,包括：

[0059] S401、根据所述将使用的报表类型，从预设的数据库中调取预设的多个图表模板与多个文字部分模板；

[0060] S402、将所述当前用户选择的图表模板与文字部分模板组合成为所述初步报表；

[0061] S403、调取所述初步报表。

[0062] 如上所述，实现了从数据库中调取预设的初步报表。其中，所述初步报表由图

表模板（包括图表样式模板和图表数据模板）与文字部分模板组合生成。由于不同报表需要的具体要求不同，因此报表的格式、布局、图表等各方面都不尽相同。通过将报表分解为图表和文字部分，再针对性地预先设计有图表模板与文字部分模板并存储于数据库中，当需要特定报表时，只需从已有的图表模板与文字部分模板中选择，再组合起来，即可形成所述初步报表。

[0063] 在一个实施方式中，所述根据所述当前用户输入的信息，调整所述初步报表，从而获得最终报表的步骤S5,包括：

[0064] S501、根据所述当前用户输入的图表调整信息、图表数据内容信息和文字部分调整信息，调整所述初步报表中的图表和文字部分；

[0065] S502、将所述当前用户输入的文字内容填入所述初步报表的文字部分，获得最终报表。

[0066] 如上所述，实现了根据所述当前用户输入的信息，调整所述初步报表，从而获得最终报表。如前述，已经得到了初步报表。但是所述初步报表的一些细节参数与具体数据内容并未补充。据此，根据当前用户的具体指令进行调整，其中，所述图表调整信息包括：调整图表的大小、图表的数据显示参数（如时间轴的单位时间长度等）等；所述图表数据内容信息包括：图表数据（如曲线图的数据点，饼状图的各区块占比等）等；所述文字部分调整信息包括：调整字体大小、颜色等。再将所述当前用户输入的文字内容填入所述文字部分，获得最终报表。

[0067] 在一个实施方式中，所述报表类型预测模型的获取方法，包括：

[0068] S201、获取包括指定量的样本数据的训练集，其中，所述样本数据包括用户特征信息，以及与所述用户特征信息对应的报表类型；

[0069] S202、将所述训练集的样本数据输入到神经网络模型中进行训练，其中，训练的过程中采用随机梯度下降法，利用反向传导法则更新所述神经网络模型各层的参数，得到初步训练模型；

[0070] S203、将所述初步训练模型记为所述报表类型预测模型。

[0071] 如上所述，实现了获取所述报表类型预测模型。其中，所述用户的特征信息是指，能够体现用户的特性的信息，例如用户的职业、当前近期（指定时间内）

使用的报表类型、用户的年龄、用户的性别。本实施方式中的机器学习，采用神经网络模型，例如采用VGG-F模型、VGG16模型、InceptionV3模型、Xception模型、AlexNet模型等，再以包括用户特征信息，以及与所述用户特征信息对应的报表类型的样本数据训练所述神经网络模型，其中样本数据越多，训练而得的预测模型越准确。当样品数据过多时，优选采用随机梯度下降法训练，即随机取样一些训练数据，替代整个训练集，从而提高训练速度。其中反向传导法则（BP）建立在梯度下降法的基础上，其实质上是一种映射关系：一个 n 输入 m 输出的BP神经网络所完成的功能是从 n 维欧氏空间向 m 维欧氏空间中一有限域的连续映射，以实现参数的更新。

[0072] 在一个实施方式中，所述将所述初步训练模型记为所述报表类型预测模型的步骤203，包括：

[0073] S2031、获取包括指定量的样本数据的验证集，其中，所述验证集的样本数据包括用户特征信息，以及与所述用户特征信息对应的报表类型；

[0074] S2032、利用所述验证集的样本数据验证所述初步训练模型；

[0075] S2033、如果验证通过，则将所述初步训练模型记为所述报表类型预测模型。

[0076] 如上所述，实现了获取所述报表类型预测模型。其中，所述用户的特征信息是指，能够体现用户的特性的信息，例如用户的职业、当前近期（指定时间内）使用的报表类型、用户的年龄、用户的性别。其中验证集的样本数据为了验证初步训练模型是为能够用于当前用户的报表类型预测，所以指定量的样本数据的验证集，优选与当前用户相关的样本数据，例如职业相同、年龄相同的样本数据。当验证通过，表明该初步训练模型可用，据此将所述初步训练模型记为所述报表类型预测模型。

[0077] 在一个实施方式中，所述报表类型预测模型的获取方法，包括：

[0078] S211、获取指定量的样本数据，其中，所述样本数据包括用户特征信息，以及与所述用户特征信息对应的报表类型；

[0079] S212、将所述训练集的样本数据输入到CHAID决策树模型中进行训练，得到初步CHAID决策树；

[0080] S213、将所述初步CHAID决策树记为所述报表类型预测模型。

[0081] 如上所述，实现了获得报表类型预测模型。决策树根据用户特征信息进行分类，以预测用户将采用的报表类型。其中所述用户的特征信息是指，能够体现用户的特性的信息，例如用户的职业、当前近期（指定时间内）使用的报表类型、用户的年龄、用户的性别。其中CHAID决策树模型指采用卡方自动交互检测法CHAID的决策树模型。在此简单介绍CHAID决策树的原理：1、合并组内对决策变量影响差别不显著的组值；2、选取卡方值最大的变量作为树分类变量；3、重复1、2步骤，至不能选取卡方值大于某值或P值不再小于某临界值，或样本小于某数。其中，所述CHAID的决策树模型的建模标准例如为树的最大层数为3层、母节点可再分的显著水平为0.05、母节点包含的最小样本数为100、子节点包含的最小样本数为50。进一步地，所述将所述初步CHAID决策树记为所述报表类型预测模型还包括：利用预先获得的由样本数据构成的验证集验证所述初步CHAID决策树；如果验证通过，则将所述初步CHAID决策树记为所述报表类型预测模型。

[0082] 在一个实施方式中，所述将所述训练集的样本数据输入到CHAID决策树模型中进行训练，得到初步CHAID决策树的步骤S212，包括：

[0083] S2121、设置所述CHAID决策树模型的建模标准参数，所述建模标准参数包括决策树的最大层数、母节点的可再分的显著水平、母节点包含的最小样本数和子节点包含的最小样本数；

[0084] S2122、将训练集的样本数据输入到采用卡方自动交互检测法建立的所述CHAID决策树模型中进行训练，得到初步CHAID决策树。

[0085] 如上所述，实现了得到初步CHAID决策树。其中设置所述CHAID决策树模型的建模标准参数，才可确定所述CHAID决策树模型。所述建模标准参数包括决策树的最大层数、母节点的可再分的显著水平、母节点包含的最小样本数和子节点包含的最小样本数，例如树的最大层数为3-5层、母节点可再分的显著水平为0.05、母节点包含的最小样本数为100-200、子节点包含的最小样本数为50-100。

[0086] 本申请的基于机器学习的报表生成方法，通过基于机器学习的报表类型预测模型，以预测当前用户将使用的报表类型，再根据所述报表类型，从数据库中调

取预设的初步报表，对初步报表进行调整后以得到最终报表，从而避免用户的繁琐操作，提升用户体验，提高报表完成效率。

[0087] 参照图2，本申请实施例提供一种基于机器学习的报表生成装置，包括：

[0088] 特征信息获取单元10，用于获取当前用户的特征信息，所述当前用户的特征信息至少包括当前用户的职业信息；

[0089] 报表类型预测模型运算单元20，用于将所述特征信息输入预设的基于机器学习的报表类型预测模型中运算，其中所述报表类型预测模型通过由用户特征信息，以及与所述用户特征信息对应的报表类型所组成的训练数据训练而成；

[0090] 报表类型预测单元30，用于输出预测的所述当前用户将使用的报表类型；

[0091] 初步报表调取单元40，用于根据所述将使用的报表类型，从数据库中调取预设的初步报表，其中所述初步报表的类型与所述将使用的报表类型相同；

[0092] 最终报表获得单元50，用于根据所述当前用户输入的信息，调整所述初步报表，从而获得最终报表。

[0093] 如上述单元10所述，获取当前用户的特征信息，所述当前用户的特征信息至少包括当前用户的职业信息。其中，所述当前用户的特征信息是指，能够体现当前用户的特性的信息，例如当前用户的职业、当前用户近期（指定时间内）使用的报表类型、当前用户的年龄、当前用户的性别。其中，当前用户的职业信息与其可能使用的报表类型有较大关系，所以所述当前用户的特征信息至少包括当前用户的职业信息，以提高报表类型预测的准确性。例如，金融行业中的股票评论师，有较高机率使用股票报表。获取当前用户的特征信息的过程包括：从所述当前用户的注册帐户信息中提取所述当前用户的特征信息。

[0094] 如上述单元20所述，将所述特征信息输入预设的基于机器学习的报表类型预测模型中运算，其中所述报表类型预测模型通过用户特征信息，及与所述用户特征信息对应的报表类型训练而成。基于机器学习的报表类型预测模型，通过不断的自我学习，提高预测的准确性，避免预测出错误的报表类型。所述报表类型预测模型可以基于任意机器学习的模型生成，例如神经网络模型、分类树模型，再经由训练数据训练而成。其中，所述报表类型预测模型通过由用户特征信息，以及与所述用户特征信息对应的报表类型所组成的训练数据训练而成。

所述用户特征信息可以包括任意能够体现用户的特性的信息，例如用户的职业、用户近期（指定时间内）使用的报表类型、用户的年龄、用户的性别。其中，所述用户指使用报表的用户。其中，所述报表类型预测模型用于根据用户特征信息，预测出用户将要使用的报表类型。

[0095] 如上述单元30所述，输出预测的所述当前用户将使用的报表类型。所述报表类型可以为任意分类方法分成的类型，例如以包括的图表种类来分为：饼状图报表（报表中包括饼状图）、曲线图报表、直方图报表；根据用途来分为：财务报表、金融报表、统计分析报表。通过所述报表类型预测模型，即可输出预测的所述当前用户将使用的报表类型。

[0096] 如上述单元40所述，根据所述将使用的报表类型，从数据库中调取预设的初步报表，其中所述初步报表的类型与所述将使用的报表类型相同。其中预设的数据库中预存有不同报表类型的初步报表，同一报表类型的初步报表可以有多种，以供用户选择。例如饼状图报表初步报表可以包括饼状图位于报表顶部、位于报表中间、位于报表末尾等三种报表；金融初步报表可以包括具有K线图的报表、具有直方图的报表、具有曲线图的报表等。这些初步报表可供用户直接选择，从而省去了用户琐碎地从零开始建立报表的过程，提高了报表形成速度。其中，所述预设的初步报表可以是已经建好的完整的报表，也可以是由当前用户从多个备选的图表模板与文字部分模板中选出需要的确切的图表模板与文字部分模板所组合而成的报表。

[0097] 如上述单元50所述，根据所述当前用户输入的信息，调整所述初步报表，从而获得最终报表。由前述，提供了初步报表，再根据所述当前用户输入的信息，即可得到最终报表。其中，所述当前用户输入的信息，包括对初步报表的参数进行调整的信息、对初步报表进行描述的文字内容信息中的至少一种。其中，所述最终报表至少包括图表部分，进一步地，还可以包括文字部分。

[0098] 在一个实施方式中，所述初步报表调取单元40,包括：

[0099] 图表模板与文字部分模板调取子单元，用于根据所述将使用的报表类型，从预设的数据库中调取预设的多个图表模板与多个文字部分模板；

[0100] 组合子单元，用于将所述当前用户选择的图表模板与文字部分模板组合成为所

述初步报表；

[0101] 调取子单元，用于调取所述初步报表。

[0102] 如上所述，实现了从数据库中调取预设的初步报表。其中，所述初步报表由图表模板（包括图表样式模板和图表数据模板）与文字部分模板组合生成。由于不同报表需要的具体要求不同，因此报表的格式、布局、图表等各方面都不尽相同。通过将报表分解为图表和文字部分，再针对性地预先设计有图表模板与文字部分模板并存储于数据库中，当需要特定报表时，只需从已有的图表模板与文字部分模板中选择，再组合起来，即可形成所述初步报表。

[0103] 在一个实施方式中，所述最终报表获得单元50,包括：

[0104] 调整子单元，用于根据所述当前用户输入的图表调整信息、图表数据内容信息和文字部分调整信息，调整所述初步报表中的图表和文字部分；

[0105] 最终报表获得子单元，用于将所述当前用户输入的文字内容填入所述初步报表的文字部分，获得最终报表。

[0106] 如上所述，实现了根据所述当前用户输入的信息，调整所述初步报表，从而获得最终报表。如前述，已经得到了初步报表。但是所述初步报表的一些细节参数与具体数据内容并未补充。据此，根据当前用户的具体指令进行调整，其中，所述图表调整信息包括：调整图表的大小、图表的数据显示参数（如时间轴的单位时间长度等）等；所述图表数据内容信息包括：图表数据（如曲线图的数据点，饼状图的各区块占比等）等；所述文字部分调整信息包括：调整字体大小、颜色等。再将所述当前用户输入的文字内容填入所述文字部分，获得最终报表。

[0107] 在一个实施方式中，所述装置包括报表类型预测模型获取单元，所述报表类型预测模型获取单元，包括：

[0108] 训练集获取子单元，用于获取包括指定量的样本数据的训练集，其中，所述样本数据包括用户特征信息，以及与所述用户特征信息对应的报表类型；

[0109] 神经网络模型训练子单元，用于将所述训练集的样本数据输入到神经网络模型中进行训练，其中，训练的过程中采用随机梯度下降法，利用反向传导法则更新所述神经网络模型各层的参数，得到初步训练模型；

[0110] 报表类型预测模型标记子单元，用于将所述初步训练模型记为所述报表类型预测模型。

[0111] 如上所述，实现了获取所述报表类型预测模型。其中，所述用户的特征信息是指，能够体现用户的特性的信息，例如用户的职业、当前近期（指定时间内）使用的报表类型、用户的年龄、用户的性别。本实施方式中的机器学习，采用神经网络模型，例如采用VGG-F模型、VGG16模型、InceptionV3模型、Xception模型、AlexNet模型等，再以包括用户特征信息，以及与所述用户特征信息对应的报表类型的样本数据训练所述神经网络模型，其中样本数据越多，训练而得的预测模型越准确。当样品数据过多时，优选采用随机梯度下降法训练，即随机取样一些训练数据，替代整个训练集，从而提高训练速度。其中反向传导法则（BP）建立在梯度下降法的基础上，其实质上是一种映射关系：一个 n 输入 m 输出的BP神经网络所完成的功能是从 n 维欧氏空间向 m 维欧氏空间中一有限域的连续映射，以实现参数的更新。

[0112] 在一个实施方式中，所述报表类型预测模型标记子单元，包括：

[0113] 验证集获取模块，用于获取包括指定量的样本数据的验证集，其中，所述验证集的样本数据包括用户特征信息，以及与所述用户特征信息对应的报表类型；

[0114] 验证模块，用于利用所述验证集的样本数据验证所述初步训练模型；

[0115] 报表类型预测模型标记模块，用于如果验证通过，则将所述初步训练模型记为所述报表类型预测模型。

[0116] 如上所述，实现了获取所述报表类型预测模型。其中，所述用户的特征信息是指，能够体现用户的特性的信息，例如用户的职业、当前近期（指定时间内）使用的报表类型、用户的年龄、用户的性别。其中验证集的样本数据为了验证初步训练模型是为能够用于当前用户的报表类型预测，所以指定量的样本数据的验证集，优选与当前用户相关的样本数据，例如职业相同、年龄相同的样本数据。当验证通过，表明该初步训练模型可用，据此将所述初步训练模型记为所述报表类型预测模型。

[0117] 在一个实施方式中，所述装置包括报表类型预测模型获取单元，所述报表类型预测模型获取单元，包括：

- [0118] 样本数据获取子单元，用于获取指定量的样本数据，其中，所述样本数据包括用户特征信息，以及与所述用户特征信息对应的报表类型；
- [0119] 决策树模型训练子单元，用于将所述训练集的样本数据输入到CHAID决策树模型中进行训练，得到初步CHAID决策树；
- [0120] 决策树标记子单元，用于将所述初步CHAID决策树记为所述报表类型预测模型。
- [0121] 如上所述，实现了获得报表类型预测模型。决策树根据用户特征信息进行分类，以预测用户将采用的报表类型。其中所述用户的特征信息是指，能够体现用户的特性的信息，例如用户的职业、当前近期（指定时间内）使用的报表类型、用户的年龄、用户的性别。其中CHAID决策树模型指采用卡方自动交互检测法CHAID的决策树模型。在此简单介绍CHAID决策树的原理：1、合并组内对决策变量影响差别不显著的组值；2、选取卡方值最大的变量作为树分类变量；3、重复1、2步骤，至不能选取卡方值大于某值或P值不再小于某临界值，或样本小于某数。其中，所述CHAID的决策树模型的建模标准例如为树的最大层数为3层、母节点可再分的显著水平为0.05、母节点包含的最小样本数为100、子节点包含的最小样本数为50。进一步地，所述将所述初步CHAID决策树记为所述报表类型预测模型还包括：利用预先获得的由样本数据构成的验证集验证所述初步CHAID决策树；如果验证通过，则将所述初步CHAID决策树记为所述报表类型预测模型。
- [0122] 在一个实施方式中，所述决策树模型训练子单元，包括：
- [0123] 建模标准参数设置模块，用于设置所述CHAID决策树模型的建模标准参数，所述建模标准参数包括决策树的最大层数、母节点的可再分的显著水平、母节点包含的最小样本数和子节点包含的最小样本数；
- [0124] 初步CHAID决策树获得模块，用于将训练集的样本数据输入到采用卡方自动交互检测法建立的所述CHAID决策树模型中进行训练，得到初步CHAID决策树。
- [0125] 如上所述，实现了得到初步CHAID决策树。其中设置所述CHAID决策树模型的建模标准参数，才可确定所述CHAID决策树模型。所述建模标准参数包括决策树的最大层数、母节点的可再分的显著水平、母节点包含的最小样本数和子

节点包含的最小样本数，例如树的最大层数为3-5层、母节点可再分的显著水平为0.05、母节点包含的最小样本数为100-200、子节点包含的最小样本数为50-100。

[0126] 本申请的基于机器学习的报表生成装置，通过基于机器学习的报表类型预测模型，以预测当前用户将使用的报表类型，再根据所述报表类型，从数据库中调取预设的初步报表，对初步报表进行调整后以得到最终报表，从而避免用户的繁琐操作，提升用户体验，提高报表完成效率。

[0127] 参照图3，本发明实施例中还提供一种计算机设备，该计算机设备可以是服务器，其内部结构可以如图所示。该计算机设备包括通过系统总线连接的处理器、存储器、网络接口和数据库。其中，该计算机设计的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统、计算机可读指令和数据库。该内存储器为非易失性存储介质中的操作系统和计算机可读指令的运行提供环境。该计算机设备的数据库用于存储基于机器学习的报表生成方法所用数据。该计算机设备的网络接口用于与外部的终端通过网络连接通信。该计算机可读指令被处理器执行时以实现一种基于机器学习的报表生成方法。

[0128] 上述处理器执行上述基于机器学习的报表生成方法，包括以下步骤：获取当前用户的特征信息，所述当前用户的特征信息至少包括当前用户的职业信息；将所述特征信息输入预设的基于机器学习的报表类型预测模型中运算，其中所述报表类型预测模型通过由用户特征信息，以及与所述用户特征信息对应的报表类型所组成的训练数据训练而成；输出预测的所述当前用户将使用的报表类型；根据所述将使用的报表类型，从数据库中调取预设的初步报表，其中所述初步报表的类型与所述将使用的报表类型相同；根据所述当前用户输入的信息，调整所述初步报表，从而获得最终报表。

[0129] 在一个实施方式中，所述根据所述将使用的报表类型，从数据库中调取预设的初步报表，其中所述初步报表的类型与所述将使用的报表类型相同的步骤，包括：根据所述将使用的报表类型，从预设的数据库中调取预设的多个图表模板与多个文字部分模板；将所述当前用户选择的图表模板与文字部分模板组合成为

所述初步报表；调取所述初步报表。

- [0130] 在一个实施方式中，所述根据所述当前用户输入的信息，调整所述初步报表，从而获得最终报表的步骤，包括：根据所述当前用户输入的图表调整信息、图表数据内容信息和文字部分调整信息，调整所述初步报表中的图表和文字部分；将所述当前用户输入的文字内容填入所述初步报表的文字部分，获得最终报表。
- [0131] 在一个实施方式中，所述报表类型预测模型的获取方法，包括：获取包括指定量的样本数据的训练集，其中，所述样本数据包括用户特征信息，以及与所述用户特征信息对应的报表类型；将所述训练集的样本数据输入到神经网络模型中进行训练，其中，训练的过程中采用随机梯度下降法，利用反向传导法则更新所述神经网络模型各层的参数，得到初步训练模型；将所述初步训练模型记为所述报表类型预测模型。
- [0132] 在一个实施方式中，所述将所述初步训练模型记为所述报表类型预测模型的步骤，包括：获取包括指定量的样本数据的验证集，其中，所述验证集的样本数据包括用户特征信息，以及与所述用户特征信息对应的报表类型；利用所述验证集的样本数据验证所述初步训练模型；如果验证通过，则将所述初步训练模型记为所述报表类型预测模型。
- [0133] 在一个实施方式中，所述报表类型预测模型的获取方法，包括：获取指定量的样本数据，其中，所述样本数据包括用户特征信息，以及与所述用户特征信息对应的报表类型；将所述训练集的样本数据输入到CHAID决策树模型中进行训练，得到初步CHAID决策树；将所述初步CHAID决策树记为所述报表类型预测模型。
- [0134] 在一个实施方式中，所述将所述训练集的样本数据输入到CHAID决策树模型中进行训练，得到初步CHAID决策树的步骤，包括：设置所述CHAID决策树模型的建模标准参数，所述建模标准参数包括决策树的最大层数、母节点的可再分的显著水平、母节点包含的最小样本数和子节点包含的最小样本数；将训练集的样本数据输入到采用卡方自动交互检测法建立的所述CHAID决策树模型中进行训练，得到初步CHAID决策树。

- [0135] 本领域技术人员可以理解，图中示出的结构，仅仅是与本申请方案相关的部分结构的框图，并不构成对本申请方案所应用于其上的计算机设备的限定。
- [0136] 本申请的计算机设备，通过基于机器学习的报表类型预测模型，以预测当前用户将使用的报表类型，再根据所述报表类型，从数据库中调取预设的初步报表，对初步报表进行调整后以得到最终报表，从而避免用户的繁琐操作，提升用户体验，提高报表完成效率。
- [0137] 本申请一实施例还提供一种计算机可读存储介质，其上存储有计算机可读指令，计算机可读指令被处理器执行时实现基于机器学习的报表生成方法，包括以下步骤：获取当前用户的特征信息，所述当前用户的特征信息至少包括当前用户的职业信息；将所述特征信息输入预设的基于机器学习的报表类型预测模型中运算，其中所述报表类型预测模型通过由用户特征信息，以及与所述用户特征信息对应的报表类型所组成的训练数据训练而成；输出预测的所述当前用户将使用的报表类型；根据所述将使用的报表类型，从数据库中调取预设的初步报表，其中所述初步报表的类型与所述将使用的报表类型相同；根据所述当前用户输入的信息，调整所述初步报表，从而获得最终报表。所述计算机可读存储介质，例如为非易失性的计算机可读存储介质，或者为易失性的计算机可读存储介质。
- [0138] 在一个实施方式中，所述根据所述将使用的报表类型，从数据库中调取预设的初步报表，其中所述初步报表的类型与所述将使用的报表类型相同的步骤,包括：根据所述将使用的报表类型，从预设的数据库中调取预设的多个图表模板与多个文字部分模板；将所述当前用户选择的图表模板与文字部分模板组合成为所述初步报表；调取所述初步报表。
- [0139] 在一个实施方式中，所述根据所述当前用户输入的信息，调整所述初步报表，从而获得最终报表的步骤,包括：根据所述当前用户输入的图表调整信息、图表数据内容信息和文字部分调整信息，调整所述初步报表中的图表和文字部分；将所述当前用户输入的文字内容填入所述初步报表的文字部分，获得最终报表。
- [0140] 在一个实施方式中，所述报表类型预测模型的获取方法，包括：获取包括指定

量的样本数据的训练集，其中，所述样本数据包括用户特征信息，以及与所述用户特征信息对应的报表类型；将所述训练集的样本数据输入到神经网络模型中进行训练，其中，训练的过程中采用随机梯度下降法，利用反向传导法则更新所述神经网络模型各层的参数，得到初步训练模型；将所述初步训练模型记为所述报表类型预测模型。

[0141] 在一个实施方式中，所述将所述初步训练模型记为所述报表类型预测模型的步骤，包括：获取包括指定量的样本数据的验证集，其中，所述验证集的样本数据包括用户特征信息，以及与所述用户特征信息对应的报表类型；利用所述验证集的样本数据验证所述初步训练模型；如果验证通过，则将所述初步训练模型记为所述报表类型预测模型。

[0142] 在一个实施方式中，所述报表类型预测模型的获取方法，包括：获取指定量的样本数据，其中，所述样本数据包括用户特征信息，以及与所述用户特征信息对应的报表类型；将所述训练集的样本数据输入到CHAID决策树模型中进行训练，得到初步CHAID决策树；将所述初步CHAID决策树记为所述报表类型预测模型。

[0143] 在一个实施方式中，所述将所述训练集的样本数据输入到CHAID决策树模型中进行训练，得到初步CHAID决策树的步骤，包括：设置所述CHAID决策树模型的建模标准参数，所述建模标准参数包括决策树的最大层数、母节点的可再分的显著水平、母节点包含的最小样本数和子节点包含的最小样本数；将训练集的样本数据输入到采用卡方自动交互检测法建立的所述CHAID决策树模型中进行训练，得到初步CHAID决策树。

[0144] 本申请的计算机可读存储介质，通过基于机器学习的报表类型预测模型，以预测当前用户将使用的报表类型，再根据所述报表类型，从数据库中调取预设的初步报表，对初步报表进行调整后以得到最终报表，从而避免用户的繁琐操作，提升用户体验，提高报表完成效率。

[0145] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机可读指令来指令相关的硬件来完成，所述的计算机可读指令可存储于一非易失性计算机可读取存储介质中，该计算机可读指令在执行时，可包

括如上述各方法的实施例的流程。其中，本申请所提供的和实施例中所使用的对存储器、存储、数据库或其它介质的任何引用，均可包括非易失性和/或易失性存储器。非易失性存储器可以包括只读存储器（ROM）、可编程ROM（PROM）、电可编程ROM（EPROM）、电可擦除可编程ROM（EEPROM）或闪存。易失性存储器可包括随机存取存储器（RAM）或者外部高速缓冲存储器。作为说明而非局限，RAM以多种形式可得，诸如静态RAM（SRAM）、动态RAM（DRAM）、同步DRAM（SDRAM）、双速据率SDRAM（SSRSDRAM）、增强型SDRAM（ESDRAM）、同步链路（Synchlink）DRAM（SLDRAM）、存储器总线（Rambus）直接RAM（RDRAM）、直接存储器总线动态RAM（DRDRAM）、以及存储器总线动态RAM（RDRAM）等。

[0146] 需要说明的是，在本文中，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、装置、物品或者方法不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、装置、物品或者方法所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括该要素的过程、装置、物品或者方法中还存在另外的相同要素。

[0147] 以上所述仅为本申请的优选实施例，并非因此限制本申请的专利范围，凡是利用本申请说明书及附图内容所作的等效结构或等效流程变换，或直接或间接运用在其他相关的技术领域，均同理包括在本申请的专利保护范围内。

权利要求书

- [权利要求 1] 一种基于机器学习的报表生成方法，其特征在于，包括：
- 获取当前用户的特征信息，所述当前用户的特征信息至少包括当前用户的职业信息；
- 将所述特征信息输入预设的基于机器学习的报表类型预测模型中运算，其中所述报表类型预测模型通过由用户特征信息，以及与所述用户特征信息对应的报表类型所组成的训练数据训练而成；
- 输出预测的所述当前用户将使用的报表类型；
- 根据所述将使用的报表类型，从数据库中调取预设的初步报表，其中所述初步报表的类型与所述将使用的报表类型相同；
- 根据所述当前用户输入的信息，调整所述初步报表，从而获得最终报表。
- [权利要求 2] 根据权利要求1所述的基于机器学习的报表生成方法，其特征在于，所述根据所述将使用的报表类型，从数据库中调取预设的初步报表，其中所述初步报表的类型与所述将使用的报表类型相同的步骤,包括：
- ：
- 根据所述将使用的报表类型，从预设的数据库中调取预设的多个图表模板与多个文字部分模板；
- 将所述当前用户选择的图表模板与文字部分模板组合成为所述初步报表；
- 调取所述初步报表。
- [权利要求 3] 根据权利要求1所述的基于机器学习的报表生成方法，其特征在于，所述根据所述当前用户输入的信息，调整所述初步报表，从而获得最终报表的步骤,包括：
- 根据所述当前用户输入的图表调整信息、图表数据内容信息和文字部分调整信息，调整所述初步报表中的图表和文字部分；
- 将所述当前用户输入的文字内容填入所述初步报表的文字部分，获得最终报表。

- [权利要求 4] 根据权利要求1所述的基于机器学习的报表生成方法，其特征在于，所述报表类型预测模型的获取方法，包括：
- 获取包括指定量的样本数据的训练集，其中，所述样本数据包括用户特征信息，以及与所述用户特征信息对应的报表类型；
- 将所述训练集的样本数据输入到神经网络模型中进行训练，其中，训练的过程中采用随机梯度下降法，利用反向传导法则更新所述神经网络模型各层的参数，得到初步训练模型；
- 将所述初步训练模型记为所述报表类型预测模型。
- [权利要求 5] 根据权利要求4所述的基于机器学习的报表生成方法，其特征在于，所述将所述初步训练模型记为所述报表类型预测模型的步骤，包括：
- 获取包括指定量的样本数据的验证集，其中，所述验证集的样本数据包括用户特征信息，以及与所述用户特征信息对应的报表类型；
- 利用所述验证集的样本数据验证所述初步训练模型；
- 如果验证通过，则将所述初步训练模型记为所述报表类型预测模型。
- [权利要求 6] 根据权利要求1所述的基于机器学习的报表生成方法，其特征在于，所述报表类型预测模型的获取方法，包括：
- 获取指定量的样本数据，其中，所述样本数据包括用户特征信息，以及与所述用户特征信息对应的报表类型；
- 将所述训练集的样本数据输入到CHAID决策树模型中进行训练，得到初步CHAID决策树；
- 将所述初步CHAID决策树记为所述报表类型预测模型。
- [权利要求 7] 根据权利要求6所述的基于机器学习的报表生成方法，其特征在于，所述将所述训练集的样本数据输入到CHAID决策树模型中进行训练，得到初步CHAID决策树的步骤，包括：
- 设置所述CHAID决策树模型的建模标准参数，所述建模标准参数包括决策树的最大层数、母节点的可再分的显著水平、母节点包含的最小样本数和子节点包含的最小样本数；
- 将训练集的样本数据输入到采用卡方自动交互检测法建立的所述CH

AID决策树模型中进行训练，得到初步CHAID决策树。

[权利要求 8]

一种基于机器学习的报表生成装置，其特征在于，包括：

特征信息获取单元，用于获取当前用户的特征信息，所述当前用户的特征信息至少包括当前用户的职业信息；

报表类型预测模型运算单元，用于将所述特征信息输入预设的基于机器学习的报表类型预测模型中运算，其中所述报表类型预测模型通过由用户特征信息，以及与所述用户特征信息对应的报表类型所组成的训练数据训练而成；

报表类型预测单元，用于输出预测的所述当前用户将使用的报表类型；

初步报表调取单元，用于根据所述将使用的报表类型，从数据库中调取预设的初步报表，其中所述初步报表的类型与所述将使用的报表类型相同；

最终报表获得单元，用于根据所述当前用户输入的信息，调整所述初步报表，从而获得最终报表。

[权利要求 9]

根据权利要求8所述的基于机器学习的报表生成装置，其特征在于，所述初步报表调取单元,包括：

图表模板与文字部分模板调取子单元，用于根据所述将使用的报表类型，从预设的数据库中调取预设的多个图表模板与多个文字部分模板；

组合子单元，用于将所述当前用户选择的图表模板与文字部分模板组合成为所述初步报表；

调取子单元，用于调取所述初步报表。

[权利要求 10]

根据权利要求8所述的基于机器学习的报表生成装置，其特征在于，所述最终报表获得单元,包括：

调整子单元，用于根据所述当前用户输入的图表调整信息、图表数据内容信息和文字部分调整信息，调整所述初步报表中的图表和文字部分；

最终报表获得子单元，用于将所述当前用户输入的文字内容填入所述初步报表的文字部分，获得最终报表。

- [权利要求 11] 根据权利要求8所述的基于机器学习的报表生成装置，其特征在于，所述装置包括报表类型预测模型获取单元，所述报表类型预测模型获取单元，包括：
- 训练集获取子单元，用于获取包括指定量的样本数据的训练集，其中，所述样本数据包括用户特征信息，以及与所述用户特征信息对应的报表类型；
- 神经网络模型训练子单元，用于将所述训练集的样本数据输入到神经网络模型中进行训练，其中，训练的过程中采用随机梯度下降法，利用反向传导法则更新所述神经网络模型各层的参数，得到初步训练模型；
- 报表类型预测模型标记子单元，用于将所述初步训练模型记为所述报表类型预测模型。

- [权利要求 12] 根据权利要求11所述的基于机器学习的报表生成装置，其特征在于，所述报表类型预测模型标记子单元，包括：
- 验证集获取模块，用于获取包括指定量的样本数据的验证集，其中，所述验证集的样本数据包括用户特征信息，以及与所述用户特征信息对应的报表类型；
- 验证模块，用于利用所述验证集的样本数据验证所述初步训练模型；
- 报表类型预测模型标记模块，用于如果验证通过，则将所述初步训练模型记为所述报表类型预测模型。

- [权利要求 13] 根据权利要求8所述的基于机器学习的报表生成装置，其特征在于，所述装置包括报表类型预测模型获取单元，所述报表类型预测模型获取单元，包括：
- 样本数据获取子单元，用于获取指定量的样本数据，其中，所述样本数据包括用户特征信息，以及与所述用户特征信息对应的报表类型；
- 决策树模型训练子单元，用于将所述训练集的样本数据输入到CHAI

D决策树模型中进行训练，得到初步CHAID决策树；
决策树标记子单元，用于将所述初步CHAID决策树记为所述报表类型预测模型。

[权利要求 14] 根据权利要求13所述的基于机器学习的报表生成装置，其特征在于，所述决策树模型训练子单元，包括：
建模标准参数设置模块，用于设置所述CHAID决策树模型的建模标准参数，所述建模标准参数包括决策树的最大层数、母节点的可再分的显著水平、母节点包含的最小样本数和子节点包含的最小样本数；
初步CHAID决策树获得模块，用于将训练集的样本数据输入到采用卡方自动交互检测法建立的所述CHAID决策树模型中进行训练，得到初步CHAID决策树。

[权利要求 15] 一种计算机设备，包括存储器和处理器，所述存储器存储有计算机可读指令，其特征在于，所述处理器执行所述计算机可读指令时实现基于机器学习的报表生成方法，所述基于机器学习的报表生成方法，包括：
获取当前用户的特征信息，所述当前用户的特征信息至少包括当前用户的职业信息；
将所述特征信息输入预设的基于机器学习的报表类型预测模型中运算，其中所述报表类型预测模型通过由用户特征信息，以及与所述用户特征信息对应的报表类型所组成的训练数据训练而成；
输出预测的所述当前用户将使用的报表类型；
根据所述将使用的报表类型，从数据库中调取预设的初步报表，其中所述初步报表的类型与所述将使用的报表类型相同；
根据所述当前用户输入的信息，调整所述初步报表，从而获得最终报表。

[权利要求 16] 根据权利要求15所述的计算机设备，其特征在于，所述根据所述将使用的报表类型，从数据库中调取预设的初步报表，其中所述初步报表的类型与所述将使用的报表类型相同的步骤,包括：

根据所述将使用的报表类型，从预设的数据库中调取预设的多个图表模板与多个文字部分模板；

将所述当前用户选择的图表模板与文字部分模板组合成为所述初步报表；

调取所述初步报表。

[权利要求 17] 根据权利要求15所述的计算机设备，其特征在于，所述根据所述当前用户输入的信息，调整所述初步报表，从而获得最终报表的步骤,包括：

根据所述当前用户输入的图表调整信息、图表数据内容信息和文字部分调整信息，调整所述初步报表中的图表和文字部分；

将所述当前用户输入的文字内容填入所述初步报表的文字部分，获得最终报表。

[权利要求 18] 一种计算机可读存储介质，其上存储有计算机可读指令，其特征在于，所述计算机可读指令被处理器执行时实现基于机器学习的报表生成方法，所述基于机器学习的报表生成方法，包括：

获取当前用户的特征信息，所述当前用户的特征信息至少包括当前用户的职业信息；

将所述特征信息输入预设的基于机器学习的报表类型预测模型中运算，其中所述报表类型预测模型通过由用户特征信息，以及与所述用户特征信息对应的报表类型所组成的训练数据训练而成；

输出预测的所述当前用户将使用的报表类型；

根据所述将使用的报表类型，从数据库中调取预设的初步报表，其中所述初步报表的类型与所述将使用的报表类型相同；

根据所述当前用户输入的信息，调整所述初步报表，从而获得最终报表。

[权利要求 19] 根据权利要求18所述的计算机可读存储介质，其特征在于，所述根据所述将使用的报表类型，从数据库中调取预设的初步报表，其中所述初步报表的类型与所述将使用的报表类型相同的步骤,包括：

根据所述将使用的报表类型，从预设的数据库中调取预设的多个图表模板与多个文字部分模板；

将所述当前用户选择的图表模板与文字部分模板组合成为所述初步报表；

调取所述初步报表。

[权利要求 20]

根据权利要求18所述的计算机可读存储介质，其特征在于，所述根据所述当前用户输入的信息，调整所述初步报表，从而获得最终报表的步骤,包括：

根据所述当前用户输入的图表调整信息、图表数据内容信息和文字部分调整信息，调整所述初步报表中的图表和文字部分；

将所述当前用户输入的文字内容填入所述初步报表的文字部分，获得最终报表。

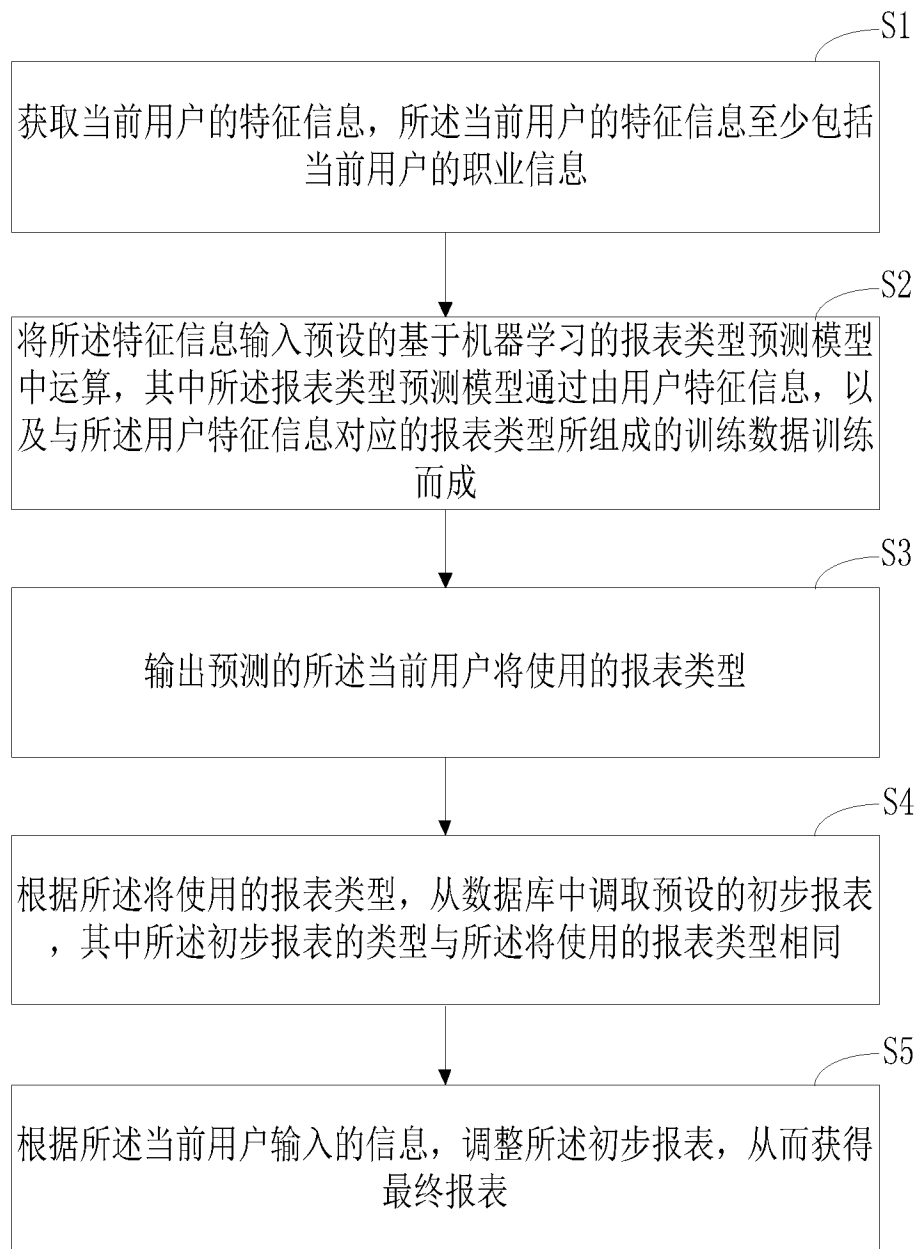


图 1

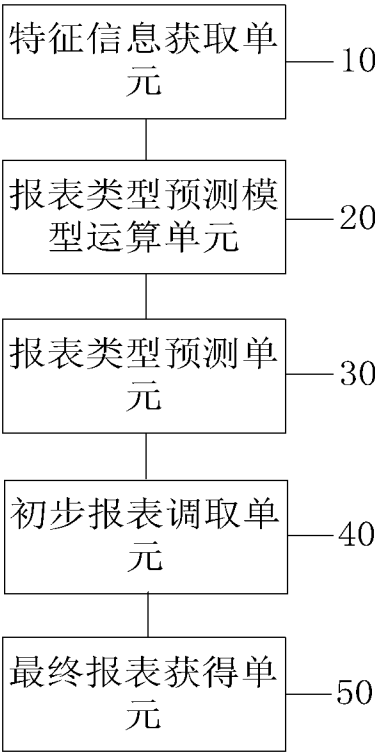


图 2

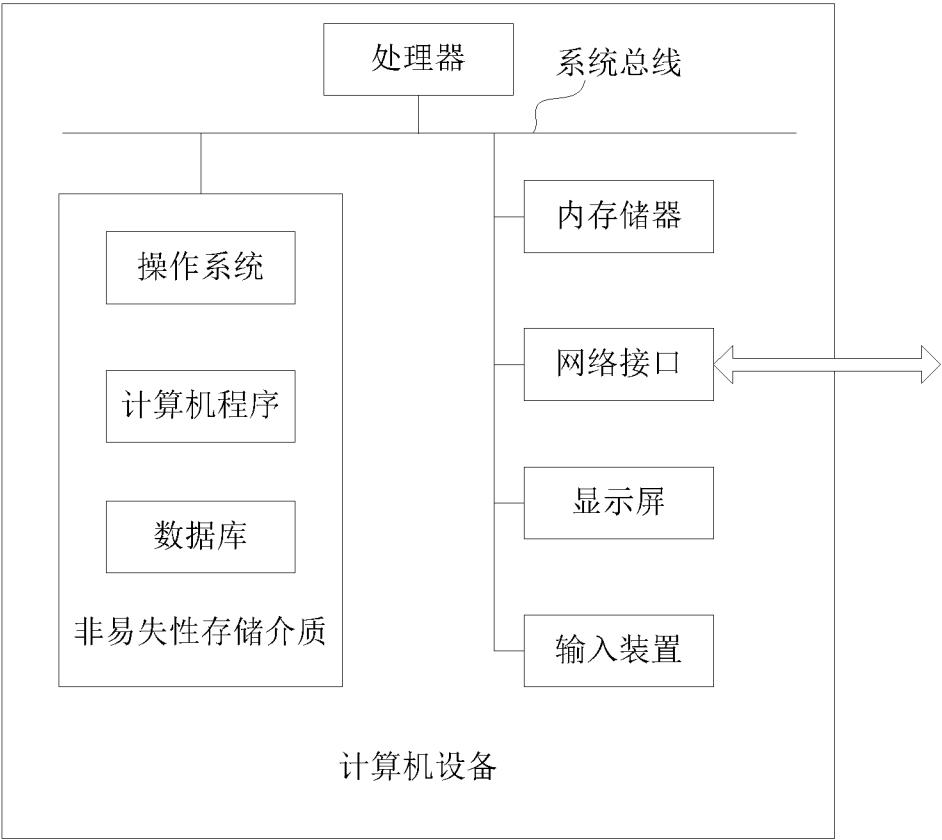


图 3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/119480

A. CLASSIFICATION OF SUBJECT MATTER

G06F 16/901(2019.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT; CNKI; WPI; EPODOC: 生成, 创建, 报表, 特征, 用户, 职业, 工作, 训练, 模型, 机器学习, generat+, establish+, report, form, character, user, job, occupation, train+, model, machine learning

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 109800333 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 24 May 2019 (2019-05-24) description, paragraphs [0046]-[0143]	1-20
X	CN 107632971 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.) 26 January 2018 (2018-01-26) description, paragraphs [0043]-[0086]	1-20
A	US 2018189655 A1 (ELECTRONICS AND TELECOMMUNICATIONS RESEARCH INSTITUTE) 05 July 2018 (2018-07-05) entire document	1-20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

13 January 2020

Date of mailing of the international search report

23 January 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/119480

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	109800333	A	24 May 2019	None			
CN	107632971	A	26 January 2018	None			
US	2018189655	A1	05 July 2018	DE	102017131259	A1	05 July 2018
				JP	2018109996	A	12 July 2018
				KR	20180080111	A	11 July 2018

国际检索报告

国际申请号

PCT/CN2019/119480

A. 主题的分类

G06F 16/901 (2019. 01) i

按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类

B. 检索领域

检索的最低限度文献 (标明分类系统和分类号)

G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用))

CNPAT; CNKI; WPI; EPDOC: 生成, 创建, 报表, 特征, 用户, 职业, 工作, 训练, 模型, 机器学习, generat+, establish+, report, form, character, user, job, occupation, train+, model, machine learning

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 109800333 A (平安科技深圳有限公司) 2019年 5月 24日 (2019 - 05 - 24) 说明书第[0046]-[0143]段	1-20
X	CN 107632971 A (百度在线网络技术北京有限公司) 2018年 1月 26日 (2018 - 01 - 26) 说明书第[0043]-[0086]段	1-20
A	US 2018189655 A1 (ELECTRONICS AND TELECOMMUNICATIONS RESEARCH INSTITUTE) 2018年 7月 5日 (2018 - 07 - 05) 全文	1-20

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2020年 1月 13日

国际检索报告邮寄日期

2020年 1月 23日

ISA/CN的名称和邮寄地址

中国国家知识产权局 (ISA/CN)

中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

授权官员

陈希元

电话号码 86-(10)-53961594

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/119480

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	109800333	A	2019年 5月 24日	无	
CN	107632971	A	2018年 1月 26日	无	
US	2018189655	A1	2018年 7月 5日	DE 102017131259 A1	2018年 7月 5日
				JP 2018109996 A	2018年 7月 12日
				KR 20180080111 A	2018年 7月 11日