

(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
2019 年 2 月 7 日 (07.02.2019)



(10) 国际公布号
WO 2019/024231 A1

- (51) 国际专利分类号:
G06F 17/30 (2006.01)
- (21) 国际申请号: PCT/CN2017/104820
- (22) 国际申请日: 2017 年 9 月 30 日 (30.09.2017)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201710660957.2 2017年8月4日 (04.08.2017) CN
- (71) 申请人: 平安科技(深圳)有限公司(PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN];
中国广东省深圳福田区八卦岭八卦三路平安大厦吴东勤, Guangdong 518000 (CN)。
- (72) 发明人: 陈娴娴(CHEN, Xianxian); 中国广东省深圳福田区八卦岭八卦三路平安大厦吴东勤,

Guangdong 518000 (CN)。李菲菲(LI, Feifei); 中国广东省深圳福田区八卦岭八卦三路平安大厦吴东勤, Guangdong 518000 (CN)。徐亮(XU, Liang); 中国广东省深圳福田区八卦岭八卦三路平安大厦吴东勤, Guangdong 518000 (CN)。肖京(XIAO, Jing); 中国广东省深圳福田区八卦岭八卦三路平安大厦吴东勤, Guangdong 518000 (CN)。

(74) 代理人: 深圳市沃德知识产权代理事务所(普通合伙)(SHENZHEN WORLD INTELLECTUAL PROPERTY AGENCY (GENERAL PARTNERSHIP)); 中国广东省深圳福田区园岭街道八卦四路10号中浩大厦1528-1530室于志光, 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,

(54) Title: AUTOMATIC DATA MATCHING METHOD, ELECTRONIC DEVICE AND COMPUTER-READABLE STORAGE MEDIUM

(54) 发明名称: 数据自动匹配方法、电子设备及计算机可读存储介质

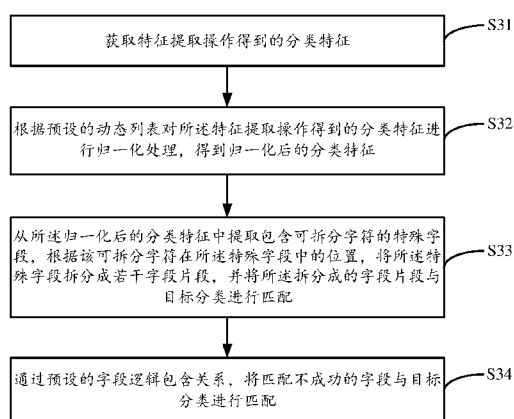


图3

- S31 ACQUIRE A CLASSIFIED FEATURE OBTAINED BY A FEATURE EXTRACTION OPERATION
- S32 PERFORM, ACCORDING TO A PRESET DYNAMIC LIST, NORMALIZATION PROCESSING OF THE CLASSIFIED FEATURE OBTAINED BY THE FEATURE EXTRACTION OPERATION, TO OBTAIN A NORMALIZED CLASSIFIED FEATURE
- S33 EXTRACT, FROM THE NORMALIZED CLASSIFIED FEATURE, A SPECIAL FIELD COMPRISING A SPLITTABLE CHARACTER, SPLIT THE SPECIAL FIELD INTO SEVERAL FIELD SEGMENTS ACCORDING TO THE POSITION OF THE SPLITTABLE CHARACTER IN THE SPECIAL FIELD, AND COMPARE THE SPLIT FIELD SEGMENTS WITH A TARGET CATEGORY
- S34 ON THE BASIS OF A PRESET FIELD LOGIC INCLUSION RELATIONSHIP, COMPARE A FIELD WHICH DOES NOT MATCH WITH THE TARGET CATEGORY

(57) Abstract: An automatic data matching method, the method comprising the following steps: acquiring a classified feature obtained by a feature extraction operation (S31); performing, according to a preset dynamic list, normalization processing of the classified feature obtained by the feature extraction operation, to obtain a normalized classified feature (S32); extracting, from the normalized classified feature, a special field comprising a splittable character, splitting the special field into several field segments according to the position of the splittable character in the special field, and comparing the split field segments with a target category (S33); and on the basis of a preset field logic inclusion relationship, comparing a field which does not match with the target category (S34). The method can improve the rate of successful matching between a classified feature and a target category and the accuracy thereof. Further disclosed are an electronic device and a computer-readable storage medium.

(57) 摘要: 一种数据自动匹配方法, 该方法包括步骤: 获取特征提取操作得到的分类特征(S31); 根据预设的动态列表对所述特征提取操作得到的分类特征进行归一化处理, 得到归一化后的分类特征(S32); 从所述归一化后的分类特征中提取包含可拆分字符的特殊字段, 根据该可拆分字符在所述特殊字段中的位置, 将所述特殊字段拆分成若干字段片段, 并将所述拆分成后的字段片段与目标分类进行匹配(S33); 通过预设的字段逻辑包含关系, 将匹配不成功的字段与目标分类进行匹配(S34)。该方法可以提升分类特征与目标分类的匹配成功率和准确率。还公开了一种电子设备及一种计算机可读存储介质。

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, 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))。

数据自动匹配方法、电子设备及计算机可读存储介质

本申请基于巴黎公约申明享有2017年8月4日递交的申请号为CN 201710660957.2、名称为“数据自动匹配方法、电子装置及计算机可读存储介质”的中国专利申请的优先权，该中国专利申请的整体内容以参考的方式结合在本申请中。

技术领域

本发明涉及计算机信息技术领域，尤其涉及一种数据自动匹配方法、电子设备及计算机可读存储介质。

背景技术

特征提取是各类数据挖掘预测模型的重要步骤，其中对分类特征按照已有的目标分类进行归一化划分在数据预处理阶段有着十分重要的作用。但是，未经清洗的分类特征与目标分类的直接的完全匹配存在着数据匹配成功率及准确率极低等问题，不能满足模型需求。并且，由于海量数据不断迁入，数据量级已远远超出手动匹配的掌控范围。故，现有技术中的数据匹配算法设计不够合理，亟需改进。

发明内容

有鉴于此，本发明提出一种数据自动匹配方法、电子设备及计算机可读存储介质，通过特殊字段结构化拆分处理和字段逻辑包含关系，有效提升了分类特征与目标分类的匹配成功率和准确率。

首先，为实现上述目的，本发明提出一种电子设备，所述电子设备包括存储器、处理器，所述存储器上存储有数据自动匹配程序，所述数据自动匹配程序被所述处理器执行时实现如下步骤：

获取特征提取操作得到的分类特征；

根据预设的动态列表对所述特征提取操作得到的分类特征进行归一化处理，得到归一化后的分类特征；

从所述归一化后的分类特征中提取包含可拆分字符的特殊字段，根据该可拆分字符在所述特殊字段中的位置，将所述特殊字段拆分成若干字段片段，并将所述拆分成的字段片段与目标分类进行匹配；及

通过预设的字段逻辑包含关系，将匹配不成功的字段与目标分类进行匹配。

优选地，所述根据预设的动态列表对所述特征提取操作得到的分类特征进行归一化处理包括：

若所述预设的动态列表为第一类型动态列表，则提取该第一类型动态列表中存储的第一类型特殊字符，根据所述提取的第一类型特殊字符对所述特征提取操作得到的分类特征进行删除或替换，得到归一化后的第一类型分类

特征；

若所述预设的动态列表为第二类型动态列表，则提取该第二类型动态列表中存储的第二类型特殊字符，根据所述提取的第二类型特殊字符对所述特征提取操作得到的分类特征进行删除或替换，得到归一化后的第二类型分类特征；及

若所述预设的动态列表为第三类型动态列表，则提取该第三类型动态列表中存储的第三类型特殊字符，根据所述提取的第三类型特殊字符对所述特征提取操作得到的分类特征进行删除或替换，得到归一化后的第三类型分类特征。

优选地，将所述特殊字段拆分成若干字段片段包括：

将该可拆分字符在所述特殊字段中的位置记录为拆分点；及

分别提取该拆分点之前的字段片段和该拆分点之后的字段片段。

优选地，将所述特殊字段拆分成若干字段片段包括：

若所述预设的动态列表为第一类型动态列表，则从所述归一化后的第一类型分类特征中提取包含可拆分字符的第一类型特殊字段，根据该可拆分字符在所述第一类型特殊字段中的位置，将所述第一类型特殊字段拆分成若干字段片段；

若所述预设的动态列表为第二类型动态列表，则从所述归一化后的第二类型分类特征中提取包含可拆分字符的第二类型特殊字段，根据该可拆分字符在所述第二类型特殊字段中的位置，将所述第二类型特殊字段拆分成若干字段片段；及

若所述预设的动态列表为第三类型动态列表，则从所述归一化后的第三类型分类特征中提取包含可拆分字符的第三类型特殊字段，根据该可拆分字符在所述第三类型特殊字段中的位置，将所述第三类型特殊字段拆分成若干字段片段。

优选地，所述通过预设的字段逻辑包含关系，将匹配不成功的字段与目标分类进行匹配包括：

根据语义逻辑相似度计算算法，计算匹配不成功的字段与目标分类的语义相似度值；及

若该语义相似度值大于预设阈值，则判定该匹配不成功的字段与该目标分类存在逻辑包含关系，将该匹配不成功的字段标记为与该目标分类存在匹配关系。

优选地，所述预设的动态列表根据数据源的数据变化进行动态调整。

优选地，所述目标分类为内部数据平台中预设的规则数据。

此外，为实现上述目的，本发明还提供一种数据自动匹配方法，该方法应用于电子设备，所述方法包括：

获取特征提取操作得到的分类特征；

根据预设的动态列表对所述特征提取操作得到的分类特征进行归一化

处理，得到归一化后的分类特征；

从所述归一化后的分类特征中提取包含可拆分字符的特殊字段，根据该可拆分字符在所述特殊字段中的位置，将所述特殊字段拆分成若干字段片段，并将所述拆分成的字段片段与目标分类进行匹配；及

通过预设的字段逻辑包含关系，将匹配不成功的字段与目标分类进行匹配。

优选地，所述根据预设的动态列表对所述特征提取操作得到的分类特征进行归一化处理包括：

若所述预设的动态列表为第一类型动态列表，则提取该第一类型动态列表中存储的第一类型特殊字符，根据所述提取的第一类型特殊字符对所述特征提取操作得到的分类特征进行删除或替换，得到归一化后的第一类型分类特征；

若所述预设的动态列表为第二类型动态列表，则提取该第二类型动态列表中存储的第二类型特殊字符，根据所述提取的第二类型特殊字符对所述特征提取操作得到的分类特征进行删除或替换，得到归一化后的第二类型分类特征；及

若所述预设的动态列表为第三类型动态列表，则提取该第三类型动态列表中存储的第三类型特殊字符，根据所述提取的第三类型特殊字符对所述特征提取操作得到的分类特征进行删除或替换，得到归一化后的第三类型分类特征。

优选地，将所述特殊字段拆分成若干字段片段包括：

将该可拆分字符在所述特殊字段中的位置记录为拆分点；及

分别提取该拆分点之前的字段片段和该拆分点之后的字段片段。

优选地，所述通过预设的字段逻辑包含关系，将匹配不成功的字段与目标分类进行匹配包括：

根据语义逻辑相似度计算算法，计算匹配不成功的字段与目标分类的语义相似度值；及

若该语义相似度值大于预设阈值，则判定该匹配不成功的字段与该目标分类存在逻辑包含关系，将该匹配不成功的字段标记为与该目标分类存在匹配关系。

优选地，所述预设的动态列表根据数据源的数据变化进行动态调整。

优选地，所述目标分类为内部数据平台中预设的规则数据。

进一步地，为实现上述目的，本发明还提供一种计算机可读存储介质，所述计算机可读存储介质存储有数据自动匹配程序，所述数据自动匹配程序可被至少一个处理器执行，以使所述至少一个处理器执行如上述的数据自动匹配方法的步骤。

相较于现有技术，本发明所提出的电子设备、数据自动匹配方法及计算

机可读存储介质，通过特殊字段结构化拆分处理，有效提升了分类特征与目标分类的匹配成功率和准确率，进一步地，通过字段逻辑包含关系（或字段语义包含关系）解决了不规则性缺失字段的匹配问题，从而进一步地提升了分类特征与目标分类的匹配成功率和准确率。

附图说明

图1是本发明电子设备一可选的硬件架构的示意图；

图2是本发明电子设备中数据自动匹配程序一实施例的模块示意图；

图3为本发明数据自动匹配方法一实施例的实施流程示意图。

附图标记：

电子设备	2
存储器	21
处理器	22
网络接口	23
数据自动匹配程序	20
获取模块	201
处理模块	202
第一匹配模块	203
第二匹配模块	204
流程步骤	S31-S34

本发明目的的实现、功能特点及优点将结合实施例，参照附图做进一步说明。

具体实施方式

为了使本发明的目的、技术方案及优点更加清楚明白，以下结合附图及实施例，对本发明进行进一步详细说明。应当理解，此处所描述的具体实施例仅用以解释本发明，并不用于限定本发明。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得的所有其他实施例，都属于本发明保护的范围。

需要说明的是，在本发明中涉及“第一”、“第二”等的描述仅用于描述目的，而不能理解为指示或暗示其相对重要性或者隐含指明所指示的技术特征的数量。由此，限定有“第一”、“第二”的特征可以明示或者隐含地包括

至少一个该特征。另外，各个实施例之间的技术方案可以相互结合，但是必须是以本领域普通技术人员能够实现为基础，当技术方案的结合出现相互矛盾或无法实现时应当认为这种技术方案的结合不存在，也不在本发明要求的保护范围之内。

进一步需要说明的是，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、方法、物品或者装置不仅包括那些要素，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、方法、物品或者装置所固有的要素。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、方法、物品或者装置中还存在另外的相同要素。

首先，本发明提出一种电子设备 2。

参阅图 1 所示，是本发明电子设备 2 一可选的硬件架构的示意图。本实施例中，所述电子设备 2 可包括，但不限于，可通过系统总线相互通信连接存储器 21、处理器 22、网络接口 23。需要指出的是，图 1 仅示出了具有组件 21-23 的电子设备 2，但是应理解的是，并不要求实施所有示出的组件，可以替代的实施更多或者更少的组件。

其中，所述电子设备 2 可以是机架式服务器、刀片式服务器、塔式服务器或机柜式服务器等计算设备，该电子设备 2 可以是独立的服务器，也可以是多个服务器所组成的服务器集群。

所述存储器 21 至少包括一种类型的可读存储介质，所述可读存储介质包括闪存、硬盘、多媒体卡、卡型存储器（例如，SD 或 DX 存储器等）、随机访问存储器（RAM）、静态随机访问存储器（SRAM）、只读存储器（ROM）、电可擦除可编程只读存储器（EEPROM）、可编程只读存储器（PROM）、磁性存储器、磁盘、光盘等。在一些实施例中，所述存储器 21 可以是所述电子设备 2 的内部存储单元，例如该电子设备 2 的硬盘或内存。在另一些实施例中，所述存储器 21 也可以是所述电子设备 2 的外部存储设备，例如该电子设备 2 上配备的插接式硬盘，智能存储卡（Smart Media Card, SMC），安全数字（Secure Digital, SD）卡，闪存卡（Flash Card）等。当然，所述存储器 21 还可以既包括所述电子设备 2 的内部存储单元也包括其外部存储设备。本实施例中，所述存储器 21 通常用于存储安装于所述电子设备 2 的操作系统和各类应用软件，例如所述数据自动匹配程序 20 等。此外，所述存储器 21 还可以用于暂时地存储已经输出或者将要输出的各类数据。

所述处理器 22 在一些实施例中可以是中央处理器（Central Processing Unit, CPU）、控制器、微控制器、微处理器、或其他数据处理芯片。该处理器 22 通常用于控制所述电子设备 2 的总体操作，例如执行与所述电子设备 2 进行数据交互或者通信相关的控制和处理等。本实施例中，所述处理器 22 用于运行所述存储器 21 中存储的程序代码或者处理数据，例如运行所述的数据自动匹配程序 20 等。

所述网络接口 23 可包括无线网络接口或有线网络接口，该网络接口 23 通常用于在所述电子设备 2 与其他电子设备之间建立通信连接。例如，所述网络接口 23 用于通过网络将所述电子设备 2 与外部数据平台相连，在所述电子设备 2 与外部数据平台之间的建立数据传输通道和通信连接。所述网络可以是企业内部网 (Intranet)、互联网 (Internet)、全球移动通讯系统 (Global System of Mobile communication, GSM)、宽带码分多址 (Wideband Code Division Multiple Access, WCDMA)、4G 网络、5G 网络、蓝牙 (Bluetooth)、Wi-Fi 等无线或有线网络。

至此，已经详细介绍了本发明各个实施例的应用环境和相关设备的硬件结构和功能。下面，将基于上述应用环境和相关设备，提出本发明的各个实施例。

参阅图 2 所示，是本发明电子设备 2 中数据自动匹配程序 20 一实施例的模块示意图。本实施例中，所述的数据自动匹配程序 20 可以被分割成一个或多个程序模块，所述一个或者多个程序模块被存储于所述存储器 21 中，并由一个或多个处理器（本实施例中为所述处理器 22）所执行，以完成本发明。例如，在图 2 中，所述的数据自动匹配程序 20 可以被分割成获取模块 201、处理模块 202、第一匹配模块 203、以及第二匹配模块 204。本发明所称的程序模块是指能够完成特定功能的一系列计算机程序指令段。以下将就各程序模块 201-204 的功能进行详细描述。

所述获取模块 201，用于获取特征提取操作得到的分类特征。其中，所述特征提取操作为各类数据挖掘预测模型的预处理步骤。优选地，在本实施例中，所述分类特征包括，但不限于，药品名称、诊断信息、医嘱信息、医疗器材、手术类型、家族史等文本数据。

所述处理模块 202，用于根据预设的动态列表对所述特征提取操作得到的分类特征进行归一化处理，得到归一化后的分类特征。

优选地，在本实施例中，所述预设的动态列表包括不同类型数据源对应的动态列表，如第一类型数据源对应的动态列表（如 MS SQL Server 数据源对应的动态列表，以下简称为“第一类型动态列表”）、第二类型数据源对应的动态列表（如 Oracle 数据源对应的动态列表，以下简称为“第二类型动态列表”）、及第三类型数据源对应的动态列表（如 MySQL 数据源对应的动态列表，以下简称为“第三类型动态列表”）等。本领域技术人员应当可以理解，在其它实施例中，所述动态列表的数量，也可以根据数据源类型的数量增加或减少。

优选地，在本实施例中，不同类型数据源对应的动态列表中存储有不同的特殊字符，用于针对不同类型的数据源进行分类特征归一化处理。例如，所述第一类型动态列表中存储有第一类型特殊字符，用于针对第一类型数据源进行分类特征归一化处理；所述第二类型动态列表中存储有第二类型特殊字符，用于针对第二类型数据源进行分类特征归一化处理；所述第三类型动

态列表中存储有第三类型特殊字符，用于针对第三类型数据源进行分类特征归一化处理。

优选地，在本实施例中，所述预设的动态列表根据数据源的数据变化进行动态调整，如增加新的特殊字符等。例如，所述第一类型动态列表根据第一类型数据源的数据变化进行动态调整，所述第二类型动态列表根据第二类型数据源的数据变化进行动态调整，所述第三类型动态列表根据第三类型数据源的数据变化进行动态调整等。

优选地，在本实施例中，所述根据预设的动态列表对所述特征提取得到的分类特征进行归一化处理包括：提取预设的动态列表中存储的特殊字符，根据所述提取的特殊字符对所述特征提取操作得到的分类特征进行删除或替换等归一化处理。

具体而言，若所述预设的动态列表为第一类型动态列表，则提取该第一类型动态列表中存储的第一类型特殊字符（如“/”和“\”等），根据所述提取的第一类型特殊字符对所述特征提取操作得到的分类特征进行删除或替换，得到归一化后的第一类型分类特征。

若所述预设的动态列表为第二类型动态列表，则提取该第二类型动态列表中存储的第二类型特殊字符，根据所述提取的第二类型特殊字符对所述特征提取操作得到的分类特征进行删除或替换，得到归一化后的第二类型分类特征。

若所述预设的动态列表为第三类型动态列表，则提取该第三类型动态列表中存储的第三类型特殊字符，根据所述提取的第三类型特殊字符对所述特征提取操作得到的分类特征进行删除或替换，得到归一化后的第三类型分类特征。

所述第一匹配模块 203，用于从所述归一化后的分类特征中提取包含可拆分字符的特殊字段，根据该可拆分字符在所述特殊字段中的位置，将所述特殊字段拆分成若干字段片段，并将所述拆分成的字段片段与目标分类进行匹配。其中，所述目标分类可以是内部数据平台（如 Hadoop 数据平台）中预设的规则数据。

优选地，在本实施例中，将所述特殊字段拆分成若干字段片段包括：将该可拆分字符在所述特殊字段中的位置记录为拆分点；分别提取该拆分点之前的字段片段和该拆分点之后的字段片段。

举例而言，若所述归一化后的分类特征中包含特殊字段“a+b”或“a/b”，其中，“+”和“/”为可拆分字符，则将所述特殊字段拆分成字段片段“a”和“b”，然后将拆分成的字段片段“a”和“b”分别与目标分类进行匹配。

由于直接将特殊字段（如“a+b”或“a/b”）与目标分类进行匹配时，很可能会导致匹配失败。但是，如果将上述特殊字段拆分成字段片段“a”和“b”后，再将拆分成的字段片段“a”和“b”分别与目标分类进行匹配

时，匹配成功率将会大大提升。因此，本发明通过第一匹配模块 203 中所述的特殊字段结构化拆分处理，能有效提升分类特征与目标分类的匹配成功率和准确率。

优选地，在本实施例中，若所述预设的动态列表为第一类型动态列表，则从所述归一化后的第一类型分类特征中提取包含可拆分字符的第一类型特殊字段，根据该可拆分字符在所述第一类型特殊字段中的位置，将所述第一类型特殊字段拆分成若干字段片段，并将所述拆分成的字段片段与目标分类进行匹配。

若所述预设的动态列表为第二类型动态列表，则从所述归一化后的第二类型分类特征中提取包含可拆分字符的第二类型特殊字段，根据该可拆分字符在所述第二类型特殊字段中的位置，将所述第二类型特殊字段拆分成若干字段片段，并将所述拆分成的字段片段与目标分类进行匹配。

若所述预设的动态列表为第三类型动态列表，则从所述归一化后的第三类型分类特征中提取包含可拆分字符的第三类型特殊字段，根据该可拆分字符在所述第三类型特殊字段中的位置，将所述第三类型特殊字段拆分成若干字段片段，并将所述拆分成的字段片段与目标分类进行匹配。

所述第二匹配模块 204，用于通过预设的字段逻辑包含关系（或字段语义包含关系），将匹配不成功的字段与目标分类进行匹配。

优选地，在本实施例中，第一匹配模块 203 中的数据匹配可以记为第一次匹配，所述第一次匹配包括：特殊字段的匹配（即将所述特殊字段拆分成字段片段与目标分类进行匹配）与非特殊字段的匹配（即将所述归一化后的分类特征中的非特殊字段与目标分类进行匹配）。进一步地，第二匹配模块 204 中的数据匹配可以记为第二次匹配，所述第二次匹配包括：将所述第一次匹配不成功的字段与目标分类进行匹配。

优选地，在本实施例中，所述通过预设的字段逻辑包含关系，将匹配不成功的字段与目标分类进行匹配包括：

根据语义逻辑相似度计算算法（如基于树状层次计算语义相似度的算法），计算匹配不成功的字段（即第一次匹配不成功的字段）与目标分类的语义相似度值；

若该语义相似度值大于预设阈值（如 80%），则判定该匹配不成功的字段与该目标分类存在逻辑包含关系，将该匹配不成功的字段标记为与该目标分类存在匹配关系，即将该匹配不成功的字段修改为匹配成功的字段。

举例而言，若第一次匹配不成功的字段包含“阿司匹林片”，而目标分类包含字段“阿司匹林”，由于“阿司匹林片”与“阿司匹林”存在语义逻辑包含关系，则将该匹配不成功的字段“阿司匹林片”修改为匹配成功的字段。

由于第二匹配模块 204 将第一匹配模块 203 中第一次匹配不成功的字段

进一步进行匹配,若发现第一次匹配不成功的字段与目标分类存在逻辑包含关系(或语义包含关系),则将第一次匹配不成功的字段修改为匹配成功的字段。因此,本发明通过第二匹配模块204中所述的字段逻辑包含关系(或字段语义包含关系)解决了不规则性缺失字段的匹配问题,从而进一步地提升了分类特征与目标分类的匹配成功率和准确率,且匹配效率与手动匹配相比有显著性优势,大大降低了人工匹配的工作量。

需要说明的是,在其它实施例中,在某些情况下,例如,所述第一次匹配成功率已经较高(如大于90%)的情形下,所述第二匹配模块204也可以去除。

通过上述程序模块201-204,本发明所提出的数据自动匹配程序20,通过特殊字段结构化拆分处理,有效提升了分类特征与目标分类的匹配成功率和准确率,进一步地,通过字段逻辑包含关系(或字段语义包含关系)解决了不规则性缺失字段的匹配问题,从而进一步地提升了分类特征与目标分类的匹配成功率和准确率。

此外,本发明还提出一种数据自动匹配方法。

参阅图3所示,是本发明数据自动匹配方法一实施例的实施流程示意图。在本实施例中,根据不同的需求,图3所示的流程图中的步骤的执行顺序可以改变,某些步骤可以省略。

步骤S31,获取特征提取操作得到的分类特征。其中,所述特征提取操作作为各类数据挖掘预测模型的预处理步骤。优选地,在本实施例中,所述分类特征包括,但不限于,药品名称、诊断信息、医嘱信息、医疗器材、手术类型、家族史等文本数据。

步骤S32,根据预设的动态列表对所述特征提取操作得到的分类特征进行归一化处理,得到归一化后的分类特征。

优选地,在本实施例中,所述预设的动态列表包括不同类型数据源对应的动态列表,如第一类型数据源对应的动态列表(如MS SQL Server数据源对应的动态列表,以下简称为“第一类型动态列表”)、第二类型数据源对应的动态列表(如Oracle数据源对应的动态列表,以下简称为“第二类型动态列表”)、及第三类型数据源对应的动态列表(如MySQL数据源对应的动态列表,以下简称为“第三类型动态列表”)等。本领域技术人员应当可以理解,在其它实施例中,所述动态列表的数量,也可以根据数据源类型的数量增加或减少。

优选地,在本实施例中,不同类型数据源对应的动态列表中存储有不同的特殊字符,用于针对不同类型的数据源进行分类特征归一化处理。例如,所述第一类型动态列表中存储有第一类型特殊字符,用于针对第一类型数据源进行分类特征归一化处理;所述第二类型动态列表中存储有第二类型特殊字符,用于针对第二类型数据源进行分类特征归一化处理;所述第三类型动

态列表中存储有第三类型特殊字符，用于针对第三类型数据源进行分类特征归一化处理。

优选地，在本实施例中，所述预设的动态列表根据数据源的数据变化进行动态调整，如增加新的特殊字符等。例如，所述第一类型动态列表根据第一类型数据源的数据变化进行动态调整，所述第二类型动态列表根据第二类型数据源的数据变化进行动态调整，所述第三类型动态列表根据第三类型数据源的数据变化进行动态调整等。

优选地，在本实施例中，所述根据预设的动态列表对所述特征提取得到的分类特征进行归一化处理包括：提取预设的动态列表中存储的特殊字符，根据所述提取的特殊字符对所述特征提取操作得到的分类特征进行删除或替换等归一化处理。

具体而言，若所述预设的动态列表为第一类型动态列表，则提取该第一类型动态列表中存储的第一类型特殊字符（如“/”和“\”等），根据所述提取的第一类型特殊字符对所述特征提取操作得到的分类特征进行删除或替换，得到归一化后的第一类型分类特征。

若所述预设的动态列表为第二类型动态列表，则提取该第二类型动态列表中存储的第二类型特殊字符，根据所述提取的第二类型特殊字符对所述特征提取操作得到的分类特征进行删除或替换，得到归一化后的第二类型分类特征。

若所述预设的动态列表为第三类型动态列表，则提取该第三类型动态列表中存储的第三类型特殊字符，根据所述提取的第三类型特殊字符对所述特征提取操作得到的分类特征进行删除或替换，得到归一化后的第三类型分类特征。

步骤S33，从所述归一化后的分类特征中提取包含可拆分字符的特殊字段，根据该可拆分字符在所述特殊字段中的位置，将所述特殊字段拆分成若干字段片段，并将所述拆分成的字段片段与目标分类进行匹配。其中，所述目标分类可以是内部数据平台（如Hadoop数据平台）中预设的规则数据。

优选地，在本实施例中，将所述特殊字段拆分成若干字段片段包括：将该可拆分字符在所述特殊字段中的位置记录为拆分点；分别提取该拆分点之前的字段片段和该拆分点之后的字段片段。

举例而言，若所述归一化后的分类特征中包含特殊字段“a+b”或“a/b”，其中，“+”和“/”为可拆分字符，则将所述特殊字段拆分成字段片段“a”和“b”，然后将拆分成的字段片段“a”和“b”分别与目标分类进行匹配。

由于直接将特殊字段（如“a+b”或“a/b”）与目标分类进行匹配时，很可能会导致匹配失败。但是，如果将上述特殊字段拆分成字段片段“a”和“b”后，再将拆分成的字段片段“a”和“b”分别与目标分类进行匹配时，匹配成功率将会大大提升。因此，本发明通过步骤S33中所述的特殊字

段结构化拆分处理，能有效提升分类特征与目标分类的匹配成功率和准确率。

优选地，在本实施例中，若所述预设的动态列表为第一类型动态列表，则从所述归一化后的第一类型分类特征中提取包含可拆分字符的第一类型特殊字段，根据该可拆分字符在所述第一类型特殊字段中的位置，将所述第一类型特殊字段拆分成若干字段片段，并将所述拆分成的字段片段与目标分类进行匹配。

若所述预设的动态列表为第二类型动态列表，则从所述归一化后的第二类型分类特征中提取包含可拆分字符的第二类型特殊字段，根据该可拆分字符在所述第二类型特殊字段中的位置，将所述第二类型特殊字段拆分成若干字段片段，并将所述拆分成的字段片段与目标分类进行匹配。

若所述预设的动态列表为第三类型动态列表，则从所述归一化后的第三类型分类特征中提取包含可拆分字符的第三类型特殊字段，根据该可拆分字符在所述第三类型特殊字段中的位置，将所述第三类型特殊字段拆分成若干字段片段，并将所述拆分成的字段片段与目标分类进行匹配。

步骤S34，通过预设的字段逻辑包含关系（或字段语义包含关系），将匹配不成功的字段与目标分类进行匹配。

优选地，在本实施例中，步骤S33中的数据匹配可以记为第一次匹配，所述第一次匹配包括：特殊字段的匹配（即将所述特殊字段拆分成字段片段与目标分类进行匹配）与非特殊字段的匹配（即将所述归一化后的分类特征中的非特殊字段与目标分类进行匹配）。进一步地，步骤S34中的数据匹配可以记为第二次匹配，所述第二次匹配包括：将所述第一次匹配不成功的字段与目标分类进行匹配。

优选地，在本实施例中，所述通过预设的字段逻辑包含关系，将匹配不成功的字段与目标分类进行匹配的步骤包括：

根据语义逻辑相似度计算算法（如基于树状层次计算语义相似度的算法），计算匹配不成功的字段（即第一次匹配不成功的字段）与目标分类的语义相似度值；

若该语义相似度值大于预设阈值（如80%），则判定该匹配不成功的字段与该目标分类存在逻辑包含关系，将该匹配不成功的字段标记为与该目标分类存在匹配关系，即将该匹配不成功的字段修改为匹配成功的字段。

举例而言，若第一次匹配不成功的字段包含“阿司匹林片”，而目标分类包含字段“阿司匹林”，由于“阿司匹林片”与“阿司匹林”存在语义逻辑包含关系，则将该匹配不成功的字段“阿司匹林片”修改为匹配成功的字段。

由于步骤S34将步骤S33中第一次匹配不成功的字段进一步进行匹配，若发现第一次匹配不成功的字段与目标分类存在逻辑包含关系（或语义包含关

系), 则将第一次匹配不成功的字段修改为匹配成功的字段。因此, 本发明通过步骤S34中所述的字段逻辑包含关系(或字段语义包含关系)解决了不规则性缺失字段的匹配问题, 从而进一步地提升了分类特征与目标分类的匹配成功率和准确率, 且匹配效率与手动匹配相比有显著性优势, 大大降低了人工匹配的工作量。

需要说明的是, 在其它实施例中, 在某些情况下, 例如, 所述第一次匹配成功率已经较高(如大于90%)的情形下, 所述步骤S34也可以去除。

通过上述步骤 S31-S34, 本发明所提出的数据自动匹配方法, 通过特殊字段结构化拆分处理, 有效提升了分类特征与目标分类的匹配成功率和准确率, 进一步地, 通过字段逻辑包含关系(或字段语义包含关系)解决了不规则性缺失字段的匹配问题, 从而进一步地提升了分类特征与目标分类的匹配成功率和准确率。

进一步地, 为实现上述目的, 本发明还提供一种计算机可读存储介质(如ROM/RAM、磁碟、光盘), 所述计算机可读存储介质存储有数据自动匹配程序20, 所述数据自动匹配程序20可被至少一个处理器22执行, 以使所述至少一个处理器22执行如上所述的数据自动匹配方法的步骤。

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上述实施例方法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来实现, 但很多情况下前者是更佳的实施方式。基于这样的理解, 本发明的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来, 该计算机软件产品存储在一个存储介质(如ROM/RAM、磁碟、光盘)中, 包括若干指令用以使得一台终端设备(可以是手机, 计算机, 服务器, 空调器, 或者网络设备等)执行本发明各个实施例所述的方法。

以上参照附图说明了本发明的优选实施例, 并非因此局限本发明的权利范围。上述本发明实施例序号仅仅为了描述, 不代表实施例的优劣。另外, 虽然在流程图中示出了逻辑顺序, 但是在某些情况下, 可以以不同于此处的顺序执行所示出或描述的步骤。

本领域技术人员不脱离本发明的范围和实质, 可以有多种变型方案实现本发明, 比如作为一个实施例的特征可用于另一实施例而得到又一实施例。凡是利用本发明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, 均同理包括在本发明的专利保护范围内。

权 利 要 求 书

1. 一种电子设备，其特征在于，所述电子设备包括存储器、处理器，所述存储器上存储有数据自动匹配程序，所述数据自动匹配程序被所述处理器执行时实现如下步骤：

获取特征提取操作得到的分类特征；

根据预设的动态列表对所述特征提取操作得到的分类特征进行归一化处理，得到归一化后的分类特征；

从所述归一化后的分类特征中提取包含可拆分字符的特殊字段，根据该可拆分字符在所述特殊字段中的位置，将所述特殊字段拆分成若干字段片段，并将所述拆分成的字段片段与目标分类进行匹配；及

通过预设的字段逻辑包含关系，将匹配不成功的字段与目标分类进行匹配。

2. 如权利要求 1 所述的电子设备，其特征在于，所述根据预设的动态列表对所述特征提取操作得到的分类特征进行归一化处理包括：

若所述预设的动态列表为第一类型动态列表，则提取该第一类型动态列表中存储的第一类型特殊字符，根据所述提取的第一类型特殊字符对所述特征提取操作得到的分类特征进行删除或替换，得到归一化后的第一类型分类特征；

若所述预设的动态列表为第二类型动态列表，则提取该第二类型动态列表中存储的第二类型特殊字符，根据所述提取的第二类型特殊字符对所述特征提取操作得到的分类特征进行删除或替换，得到归一化后的第二类型分类特征；及

若所述预设的动态列表为第三类型动态列表，则提取该第三类型动态列表中存储的第三类型特殊字符，根据所述提取的第三类型特殊字符对所述特征提取操作得到的分类特征进行删除或替换，得到归一化后的第三类型分类特征。

3. 如权利要求 2 所述的电子设备，其特征在于，将所述特殊字段拆分成若干字段片段包括：

将该可拆分字符在所述特殊字段中的位置记录为拆分点；及

分别提取该拆分点之前的字段片段和该拆分点之后的字段片段。

4. 如权利要求 3 所述的电子设备，其特征在于，将所述特殊字段拆分成若干字段片段包括：

若所述预设的动态列表为第一类型动态列表，则从所述归一化后的第一类型分类特征中提取包含可拆分字符的第一类型特殊字段，根据该可拆分字符在所述第一类型特殊字段中的位置，将所述第一类型特殊字段拆分成若干

字段片段；

若所述预设的动态列表为第二类型动态列表，则从所述归一化后的第二类型分类特征中提取包含可拆分字符的第二类型特殊字段，根据该可拆分字符在所述第二类型特殊字段中的位置，将所述第二类型特殊字段拆分成若干字段片段；及

若所述预设的动态列表为第三类型动态列表，则从所述归一化后的第三类型分类特征中提取包含可拆分字符的第三类型特殊字段，根据该可拆分字符在所述第三类型特殊字段中的位置，将所述第三类型特殊字段拆分成若干字段片段。

5. 如权利要求 1 所述的电子设备，其特征在于，所述通过预设的字段逻辑包含关系，将匹配不成功的字段与目标分类进行匹配包括：

根据语义逻辑相似度计算算法，计算匹配不成功的字段与目标分类的语义相似度值；及

若该语义相似度值大于预设阈值，则判定该匹配不成功的字段与该目标分类存在逻辑包含关系，将该匹配不成功的字段标记为与该目标分类存在匹配关系。

6. 如权利要求 1 所述的电子设备，其特征在于，所述预设的动态列表根据数据源的数据变化进行动态调整。

7. 如权利要求 1 所述的电子设备，其特征在于，所述目标分类为内部数据平台中预设的规则数据。

8. 一种数据自动匹配方法，应用于电子设备，其特征在于，所述方法包括：

获取特征提取操作得到的分类特征；

根据预设的动态列表对所述特征提取操作得到的分类特征进行归一化处理，得到归一化后的分类特征；

从所述归一化后的分类特征中提取包含可拆分字符的特殊字段，根据该可拆分字符在所述特殊字段中的位置，将所述特殊字段拆分成若干字段片段，并将所述拆分成的字段片段与目标分类进行匹配；及

通过预设的字段逻辑包含关系，将匹配不成功的字段与目标分类进行匹配。

9. 如权利要求 8 所述的数据自动匹配方法，其特征在于，所述根据预设的动态列表对所述特征提取操作得到的分类特征进行归一化处理包括：

若所述预设的动态列表为第一类型动态列表，则提取该第一类型动态列表中存储的第一类型特殊字符，根据所述提取的第一类型特殊字符对所述特

征提取操作得到的分类特征进行删除或替换,得到归一化后的第一类型分类特征;

若所述预设的动态列表为第二类型动态列表,则提取该第二类型动态列表中存储的第二类型特殊字符,根据所述提取的第二类型特殊字符对所述特征提取操作得到的分类特征进行删除或替换,得到归一化后的第二类型分类特征;及

若所述预设的动态列表为第三类型动态列表,则提取该第三类型动态列表中存储的第三类型特殊字符,根据所述提取的第三类型特殊字符对所述特征提取操作得到的分类特征进行删除或替换,得到归一化后的第三类型分类特征。

10. 如权利要求 8 所述的数据自动匹配方法,其特征在于,将所述特殊字段拆分成若干字段片段包括:

将该可拆分字符在所述特殊字段中的位置记录为拆分点;及

分别提取该拆分点之前的字段片段和该拆分点之后的字段片段。

11. 如权利要求 10 所述的数据自动匹配方法,其特征在于,将所述特殊字段拆分成若干字段片段包括:

若所述预设的动态列表为第一类型动态列表,则从所述归一化后的第一类型分类特征中提取包含可拆分字符的第一类型特殊字段,根据该可拆分字符在所述第一类型特殊字段中的位置,将所述第一类型特殊字段拆分成若干字段片段;

若所述预设的动态列表为第二类型动态列表,则从所述归一化后的第二类型分类特征中提取包含可拆分字符的第二类型特殊字段,根据该可拆分字符在所述第二类型特殊字段中的位置,将所述第二类型特殊字段拆分成若干字段片段;及

若所述预设的动态列表为第三类型动态列表,则从所述归一化后的第三类型分类特征中提取包含可拆分字符的第三类型特殊字段,根据该可拆分字符在所述第三类型特殊字段中的位置,将所述第三类型特殊字段拆分成若干字段片段。

12. 如权利要求 8 所述的数据自动匹配方法,其特征在于,所述通过预设的字段逻辑包含关系,将匹配不成功的字段与目标分类进行匹配包括:

根据语义逻辑相似度计算算法,计算匹配不成功的字段与目标分类的语义相似度值;及

若该语义相似度值大于预设阈值,则判定该匹配不成功的字段与该目标分类存在逻辑包含关系,将该匹配不成功的字段标记为与该目标分类存在匹配关系。

13. 如权利要求 8 所述的数据自动匹配方法，其特征在于，所述预设的动态列表根据数据源的数据变化进行动态调整。

14. 如权利要求 8 所述的数据自动匹配方法，其特征在于，所述目标分类为内部数据平台中预设的规则数据。

15. 一种计算机可读存储介质，所述计算机可读存储介质存储有数据自动匹配程序，所述数据自动匹配程序可被至少一个处理器执行，以使所述至少一个处理器执行以下步骤：

获取特征提取操作得到的分类特征；

根据预设的动态列表对所述特征提取操作得到的分类特征进行归一化处理，得到归一化后的分类特征；

从所述归一化后的分类特征中提取包含可拆分字符的特殊字段，根据该可拆分字符在所述特殊字段中的位置，将所述特殊字段拆分成若干字段片段，并将所述拆分成的字段片段与目标分类进行匹配；及

通过预设的字段逻辑包含关系，将匹配不成功的字段与目标分类进行匹配。

16. 如权利要求 15 所述的计算机可读存储介质，其特征在于，所述根据预设的动态列表对所述特征提取操作得到的分类特征进行归一化处理包括：

若所述预设的动态列表为第一类型动态列表，则提取该第一类型动态列表中存储的第一类型特殊字符，根据所述提取的第一类型特殊字符对所述特征提取操作得到的分类特征进行删除或替换，得到归一化后的第一类型分类特征；

若所述预设的动态列表为第二类型动态列表，则提取该第二类型动态列表中存储的第二类型特殊字符，根据所述提取的第二类型特殊字符对所述特征提取操作得到的分类特征进行删除或替换，得到归一化后的第二类型分类特征；及

若所述预设的动态列表为第三类型动态列表，则提取该第三类型动态列表中存储的第三类型特殊字符，根据所述提取的第三类型特殊字符对所述特征提取操作得到的分类特征进行删除或替换，得到归一化后的第三类型分类特征。

17. 如权利要求 15 所述的计算机可读存储介质，其特征在于，将所述特殊字段拆分成若干字段片段包括：

将该可拆分字符在所述特殊字段中的位置记录为拆分点；及

分别提取该拆分点之前的字段片段和该拆分点之后的字段片段。

18. 如权利要求 17 所述的计算机可读存储介质，其特征在于，将所述特

殊字段拆分成若干字段片段包括：

若所述预设的动态列表为第一类型动态列表，则从所述归一化后的第一类型分类特征中提取包含可拆分字符的第一类型特殊字段，根据该可拆分字符在所述第一类型特殊字段中的位置，将所述第一类型特殊字段拆分成若干字段片段；

若所述预设的动态列表为第二类型动态列表，则从所述归一化后的第二类型分类特征中提取包含可拆分字符的第二类型特殊字段，根据该可拆分字符在所述第二类型特殊字段中的位置，将所述第二类型特殊字段拆分成若干字段片段；及

若所述预设的动态列表为第三类型动态列表，则从所述归一化后的第三类型分类特征中提取包含可拆分字符的第三类型特殊字段，根据该可拆分字符在所述第三类型特殊字段中的位置，将所述第三类型特殊字段拆分成若干字段片段。

19. 如权利要求 15 所述的计算机可读存储介质，其特征在于，所述通过预设的字段逻辑包含关系，将匹配不成功的字段与目标分类进行匹配包括：

根据语义逻辑相似度计算算法，计算匹配不成功的字段与目标分类的语义相似度值；及

若该语义相似度值大于预设阈值，则判定该匹配不成功的字段与该目标分类存在逻辑包含关系，将该匹配不成功的字段标记为与该目标分类存在匹配关系。

20. 如权利要求 15 所述的计算机可读存储介质，其特征在于，所述预设的动态列表根据数据源的数据变化进行动态调整。

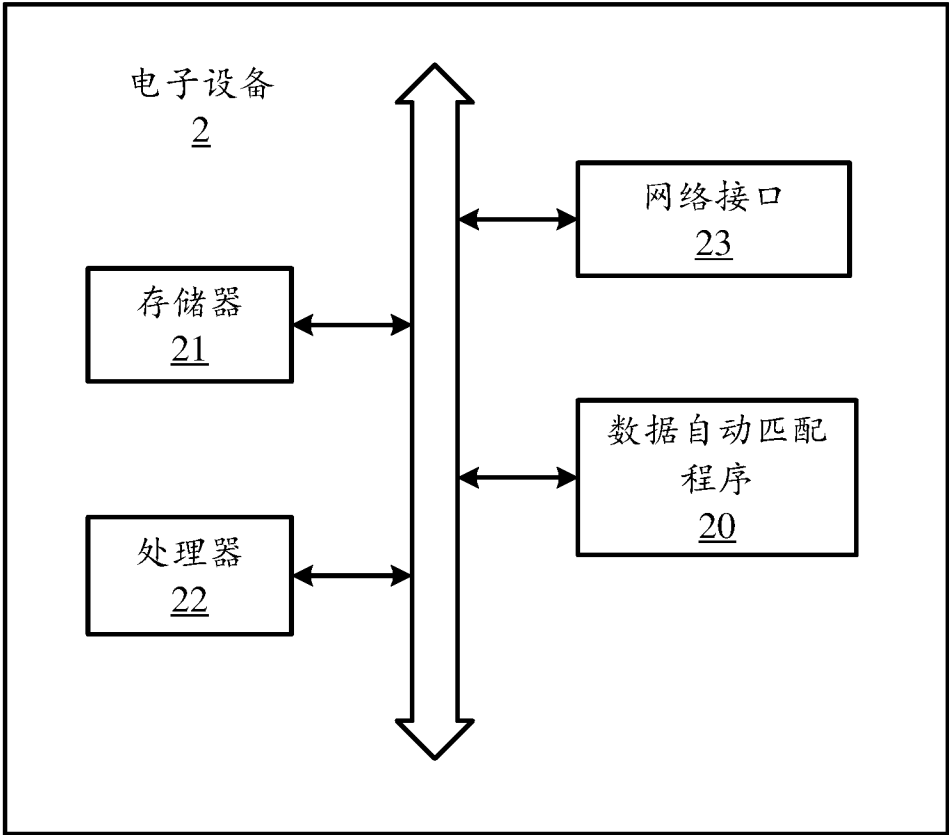


图 1

2/3

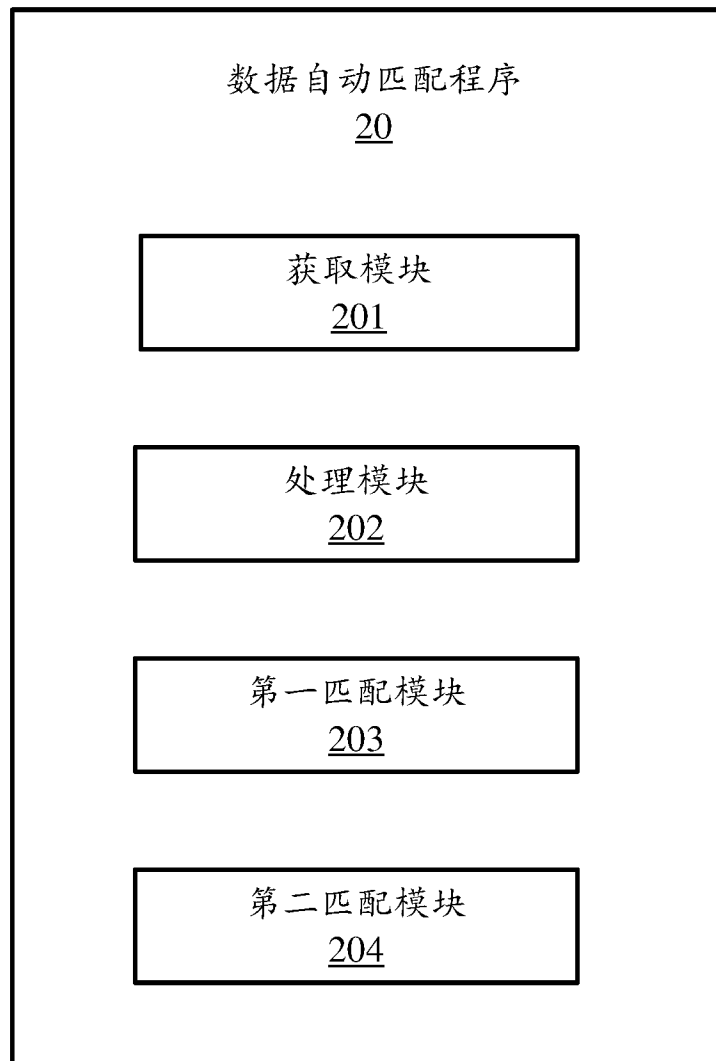


图 2

3/3

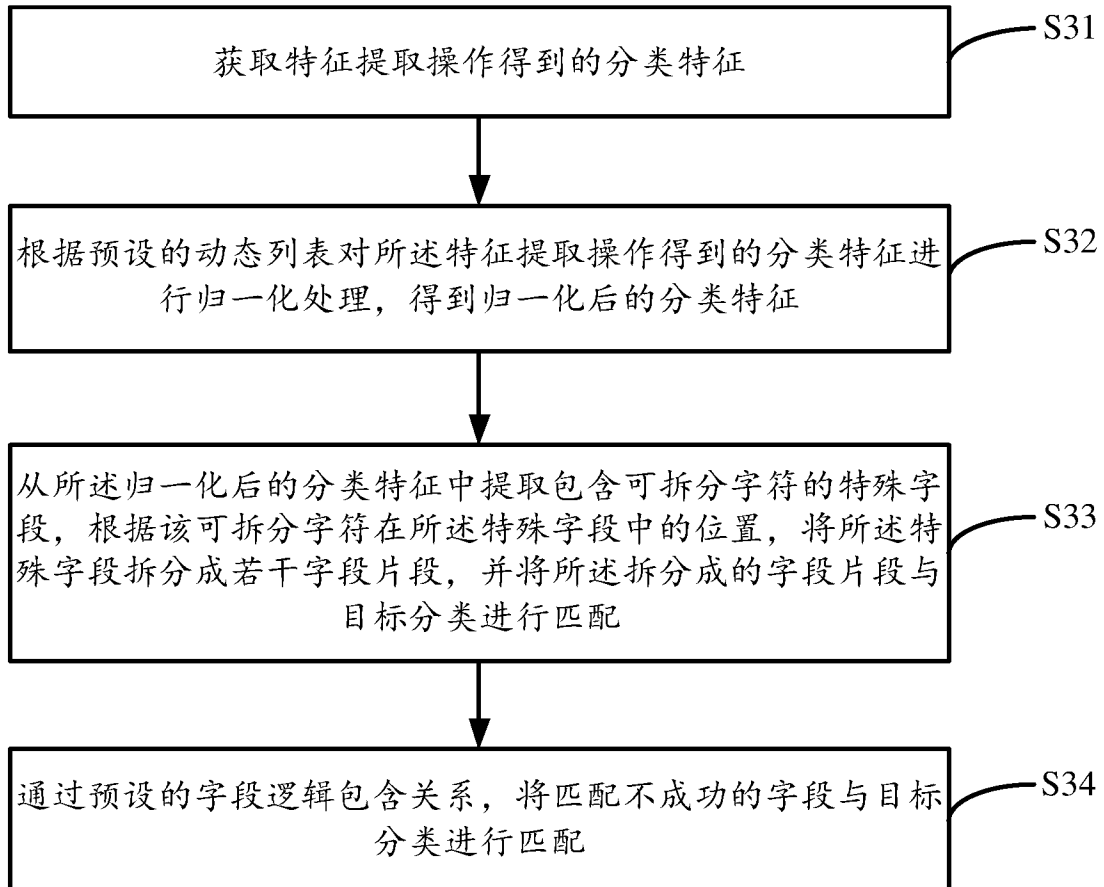


图 3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/104820

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/30 (2006.01) 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)

WPI, EPODOC, CNKI, CNPAT: 数据, 记录, 字段, 类别, 分类, 特征, 处理, 拆分, 分, 匹配, data, record, field, sort, character, feature, deal, split, match

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 107679544 A (PING'AN TECHNOLOGY (SHENZHEN) CO., LTD.) 09 February 2018 (09.02.2018), claims 1-10, and description, page 14, paragraphs 1-6	1-20
X	CN 104731976 A (MASSIVE CLOUD (BEIJING) DATA TECHNOLOGY CO., LTD.) 24 June 2015 (24.06.2015), description, paragraphs [0011]-[0051]	1-20
A	CN 106649890 A (SHUIYUN NETWORK TECHNOLOGY SERVICE CO., LTD.) 10 May 2017 (10.05.2017), entire document	1-20
A	CN 102436475 A (UFIDA SOFTWARE CO., LTD.) 02 May 2012 (02.05.2012), entire document	1-20

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	"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
"A" document defining the general state of the art which is not considered to be of particular relevance	"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
"E" earlier application or patent but published on or after the international filing date	"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
"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)	"&" document member of the same patent family
"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	

Date of the actual completion of the international search 26 March 2018	Date of mailing of the international search report 26 April 2018
Name and mailing address of the ISA State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No. (86-10) 62019451	Authorized officer HU, Ping Telephone No. (86-10) 53961335

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2017/104820

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 107679544 A	09 February 2018	None	
CN 104731976 A	24 June 2015	None	
CN 106649890 A	10 May 2017	None	
CN 102436475 A	02 May 2012	None	

国际检索报告

国际申请号

PCT/CN2017/104820

A. 主题的分类 G06F 17/30 (2006.01) 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)) WPI, EPDOC, CNKI, CNPAT: 数据, 记录, 字段, 类别, 分类, 特征, 处理, 拆分, 分, 匹配, data, record, field, sort, character, feature, deal, split, match																	
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>PX</td> <td>CN 107679544 A (平安科技深圳有限公司) 2018年 2月 9日 (2018 - 02 - 09) 权利要求1-10、说明书第14页第1-6段</td> <td>1-20</td> </tr> <tr> <td>X</td> <td>CN 104731976 A (海量云图北京数据技术有限公司) 2015年 6月 24日 (2015 - 06 - 24) 说明书第[0011]-[0051]段</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 106649890 A (税云网络科技服务有限公司) 2017年 5月 10日 (2017 - 05 - 10) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 102436475 A (用友软件股份有限公司) 2012年 5月 2日 (2012 - 05 - 02) 全文</td> <td>1-20</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	PX	CN 107679544 A (平安科技深圳有限公司) 2018年 2月 9日 (2018 - 02 - 09) 权利要求1-10、说明书第14页第1-6段	1-20	X	CN 104731976 A (海量云图北京数据技术有限公司) 2015年 6月 24日 (2015 - 06 - 24) 说明书第[0011]-[0051]段	1-20	A	CN 106649890 A (税云网络科技服务有限公司) 2017年 5月 10日 (2017 - 05 - 10) 全文	1-20	A	CN 102436475 A (用友软件股份有限公司) 2012年 5月 2日 (2012 - 05 - 02) 全文	1-20
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求															
PX	CN 107679544 A (平安科技深圳有限公司) 2018年 2月 9日 (2018 - 02 - 09) 权利要求1-10、说明书第14页第1-6段	1-20															
X	CN 104731976 A (海量云图北京数据技术有限公司) 2015年 6月 24日 (2015 - 06 - 24) 说明书第[0011]-[0051]段	1-20															
A	CN 106649890 A (税云网络科技服务有限公司) 2017年 5月 10日 (2017 - 05 - 10) 全文	1-20															
A	CN 102436475 A (用友软件股份有限公司) 2012年 5月 2日 (2012 - 05 - 02) 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 2018年 3月 26日		国际检索报告邮寄日期 2018年 4月 26日															
ISA/CN的名称和邮寄地址 中华人民共和国国家知识产权局 (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 胡平 电话号码 (86-10) 53961335															

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2017/104820

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	107679544	A	2018年 2月 9日	无	
CN	104731976	A	2015年 6月 24日	无	
CN	106649890	A	2017年 5月 10日	无	
CN	102436475	A	2012年 5月 2日	无	