

(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 10 月 10 日 (10.10.2019)



(10) 国际公布号
WO 2019/192136 A1

(51) 国际专利分类号:
G06Q 10/00 (2012.01)

(21) 国际申请号: PCT/CN2018/102226

(22) 国际申请日: 2018 年 8 月 24 日 (24.08.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201810298032.2 2018 年 4 月 3 日 (03.04.2018) 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: 李正洋(LI, ZhengYang); 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)。 毛小豪(MAO, XiaoHao); 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)); 中国广东省深圳市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:** ELECTRONIC DEVICE, FINANCIAL DATA PROCESSING METHOD AND SYSTEM, AND COMPUTER-READABLE STORAGE MEDIUM

(54) 发明名称: 电子装置、金融数据处理方法、系统和计算机可读存储介质

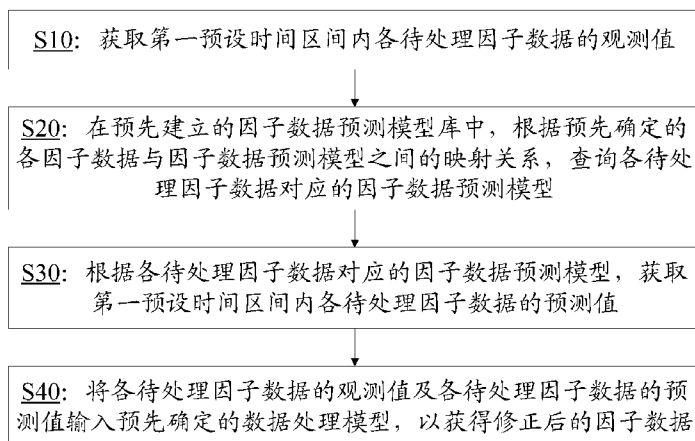


图 1

S10 Obtain an observed value of each piece of factor data to be processed within a first preset time interval
S20 According to a predetermined mapping relation between each piece of factor data and a factor data prediction model, query, from a pre-established factor data prediction model library, a factor data prediction model corresponding to the each piece of factor data to be processed
S30 Obtain, according to the factor data prediction model corresponding to the each piece of factor data to be processed, a predicted value of the each piece of factor data to be processed within the first preset time interval
S40 Input the observed value of the each piece of factor data to be processed and the predicted value of the each piece of factor data to be processed into a predetermined data processing model, to obtain corrected factor data

(57) **Abstract:** Disclosed in the present application are an electronic device, a financial data processing method and system, and a computer-readable storage medium. According to the present application, a factor data prediction model corresponding to factor data to be processed is determined, a predicted value of said factor data is calculated according to the factor data prediction model, and the observed value and predicted value of the factor data within a first preset time interval are input to a data processing model to obtain denoised factor data. Compared with the prior art, the present application eliminates the noise in the observed value of the factor data, improves the accuracy of the factor data, and lays a good foundation for the subsequent analysis of the factor data.

(57) **摘要:** 本申请公开一种电子装置、金融数据处理方法、系统和计算机可读存储介质。本申请确定待处理因子数据对应的因子数据预测模型, 根据该因子数据预测模型计算出待处理因子数据的预测值, 将第一预设时间区间内的因子数据的观测值及预测值输入至数据处理模型, 从而得到除噪后的因子数据。相较于现有技术, 本申请消除了因子数据观测值中的噪声, 提高了因子数据的准确性, 为后续针对因子数据的分析奠定了良好基础。

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))。

电子装置、金融数据处理方法、系统和计算机可读存储介质

优先权申明

本申请基于巴黎公约申明享有2018年04月03日递交的申请号为CN 201810298032.2、名称为“电子装置、金融数据处理方法和计算机可读存储介质”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

本申请涉及计算机技术领域，特别涉及一种电子装置、金融数据处理方法、系统和计算机可读存储介质。

背景技术

金融数据包括股票因子数据、债券因子数据、期货因子数据、基金因子数据等。例如，基于股票因子数据的因子分析通常被用来作为判断该股票未来走势的重要依据，其因子数据的种类多种多样，不同种类的因子数据，其相对应的计算方式也不同。由于收集方式或时空性的原因，所能观测到的因子数据基本认为是不干净（即充斥噪声）的数据。例如，由于存在延时性，在某时刻观测到的动量因子为正值时，并不代表该股票价格会继续向上增长，也可能存在动量反转现象。因此基于充斥噪声的数据进行分析得到的分析结果准确度不高。

发明内容

本申请的主要目的是解决观测到的因子数据充斥噪声，基于该因子数据进行的因子分析结果准确度不高的问题。

为实现上述目的，本申请提出的电子装置，所述电子装置包括存储器和处理器，所述存储器上存储有可在所述处理器上运行的金融数据处理系统，所述金融数据处理系统被所述处理器执行时实现如下步骤：

S10，获取第一预设时间区间内各待处理因子数据的观测值；

S20，在预先建立的因子数据预测模型库中，根据预先确定的各因子数据与因子数据预测模型之间的映射关系，查询各所述待处理因子数据对应的因子数据预测模型；

S30，根据各所述待处理因子数据对应的所述因子数据预测模型，获取所述第一预设时间区间内各所述待处理因子数据的预测值；

S40，将获取的各所述待处理因子数据的观测值及各所述待处理因子数据的预测值输入预先确定的数据处理模型，以获得修正后的因子数据。

此外，为实现上述目的，本申请提供一种金融数据处理方法，该方法包括步骤：

S10，获取第一预设时间区间内各待处理因子数据的观测值；

S20，在预先建立的因子数据预测模型库中，根据预先确定的各因子数据

与因子数据预测模型之间的映射关系，查询各所述待处理因子数据对应的因子数据预测模型；

S30，根据各所述待处理因子数据对应的所述因子数据预测模型，获取所述第一预设时间区间内各所述待处理因子数据的预测值；

S40，将获取的各所述待处理因子数据的观测值及各所述待处理因子数据的预测值输入预先确定的数据处理模型，以获得修正后的因子数据。

此外，为实现上述目的，本申请提供一种金融数据处理系统，所述金融数据处理系统包括：

第一获取模块，用于获取第一预设时间区间内各待处理因子数据的观测值；

查询模块，用于在预先建立的因子数据预测模型库中，根据预先确定的各因子数据与因子数据预测模型之间的映射关系，查询各所述待处理因子数据对应的因子数据预测模型；

第二获取模块，用于根据各所述待处理因子数据对应的所述因子数据预测模型，获取所述第一预设时间区间内各所述待处理因子数据的预测值；

数据处理模块，用于将获取的各所述待处理因子数据的观测值及各所述待处理因子数据的预测值输入预先确定的数据处理模型，以获得修正后的因子数据。

此外，为实现上述目的，本申请提供一种计算机可读存储介质，其特征在于，所述计算机可读存储介质存储有金融数据处理系统，所述金融数据处理系统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下步骤：

S10，获取第一预设时间区间内各待处理因子数据的观测值；

S20，在预先建立的因子数据预测模型库中，根据预先确定的各因子数据与因子数据预测模型之间的映射关系，查询各所述待处理因子数据对应的因子数据预测模型；

S30，根据各所述待处理因子数据对应的所述因子数据预测模型，获取所述第一预设时间区间内各所述待处理因子数据的预测值；

S40，将获取的各所述待处理因子数据的观测值及各所述待处理因子数据的预测值输入预先确定的数据处理模型，以获得修正后的因子数据。

本申请确定待处理因子数据对应的因子数据预测模型，根据该因子数据预测模型计算出待处理因子数据的预测值，将第一预设时间区间内的因子数据的观测值及预测值输入至数据处理模型，从而得到除噪后的因子数据；相较于现有技术，本申请充分考虑了各因子数据之间特性不同，针对其采用不同的因子数据预测模型可使因子数据的预测值更为准确、更为符合真实情况，此外，经数据处理模型除噪后的因子数据消除了因子数据观测值中的噪声，提高了因子数据的准确性，为后续针对因子数据的分析奠定了良好基础。

附图说明

为了更清楚地说明本申请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简单地介绍，显而易见地，下面描述中的附图仅仅是本申请的一些实施例，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示出的结构获得其他的附图。

图1为本申请金融数据处理方法第一实施例的流程示意图；

图2为本申请金融数据处理方法第二实施例的流程示意图；

图3为本申请金融数据系统第一实施例的运行环境示意图；

图4为本申请金融数据系统第一实施例的程序模块图；

图5为本申请金融数据系统第二实施例的程序模块图。

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。

具体实施方式

以下结合附图对本申请的原理和特征进行描述，所举实例只用于解释本申请，并非用于限定本申请的范围。

如图1所示，图1为本申请金融数据处理方法第一实施例的流程示意图。

本实施例中，该方法包括：

S10，获取第一预设时间区间内各待处理因子数据的观测值；

若以股票的因子数据为例，上述因子数据可包括股票当日涨幅、移动平均价（moving average）、股票交易量等。

用户可根据需要设置第一预设时间区间，例如，设置第一预设时间区间为2018年1月2日至2018年1月22日。

上述获取第一预设时间区间内各待处理因子数据的观测值的频率可根据需要设置，例如，实时，或者设置固定的时间间隔（例如，1天），或者在接收到用户发出的获取指令时执行获取第一预设时间区间内各待处理因子数据的观测值的步骤。需要注意的是，一些因子数据（例如，净利润增长、营业利润增长等）若需要通过计算得到，则可预先执行计算步骤，并存储该计算步骤得到的因子数据至预设的存储空间，当执行步骤S10时，在存储空间读取所需的因子数据即可，在一些实施例中，上述步骤S10也可包含因子数据的计算步骤。

S20，在预先建立的因子数据预测模型库中，根据预先确定的各因子数据与因子数据预测模型之间的映射关系，查询各所述待处理因子数据对应的因子数据预测模型；

预先针对各因子数据建立对应的因子数据预测模型，由于各因子数据的特性不同，针对其建立对应的因子数据预测模型的方法也不同，将建立的各因子数据预测模型存储至因子数据预测模型库中，同时保存因子数据与因子数据预测模型之间的映射关系（例如，反映因子数据与因子数据预测模型映射关系的映射表）。

S30, 根据各所述待处理因子数据对应的所述因子数据预测模型, 获取所述第一预设时间区间内各所述待处理因子数据的预测值;

例如, 若要获得 K 时刻的待处理因子数据的预测值, 则将 K-1 时刻的待处理因子数据的观测值输入至因子数据预测模型中, 因子数据预测模型输出结果即为 K 时刻的待处理因子数据的预测值。

S40, 将获取的各所述待处理因子数据的观测值及各所述待处理因子数据的预测值输入预先确定的数据处理模型, 以获得修正后的因子数据。

本实施例确定待处理因子数据对应的因子数据预测模型, 根据该因子数据预测模型计算出待处理因子数据的预测值, 将第一预设时间区间内的因子数据的观测值及预测值输入至数据处理模型, 从而得到除噪后的因子数据; 相较于现有技术, 本实施例充分考虑了各因子数据之间特性不同, 针对其采用不同的因子数据预测模型可使因子数据的预测值更为准确、更为符合真实情况, 此外, 经数据处理模型除噪后的因子数据消除了因子数据观测值中的噪声, 提高了因子数据的准确性, 为后续针对因子数据的分析奠定了良好基础。

优选的, 本实施例中, 上述步骤 S40 可具体为:

将获取的各所述待处理因子数据的观测值按照时间先后顺序进行排序, 以生成所述待处理因子数据的观测矩阵;

将获取的各所述待处理因子数据的预测值按照时间先后顺序进行排序, 以生成所述待处理因子数据的预测矩阵;

将所述待处理因子数据的观测矩阵及所述待处理因子数据的预测矩阵输入至预先确定的卡尔曼滤波模型, 以获得修正后的所述因子数据。

本实施例中, 上述生成待处理因子数据的观测矩阵的步骤还可在执行完步骤 S10 后立即执行; 同样, 上述生成待处理因子数据的预测矩阵的步骤可在执行完步骤 S30 后立即执行。

本实施例中数据处理模型为卡尔曼滤波模型, 卡尔曼滤波模型可高效的消除因子数据的噪声, 使因子数据更为接近其真实值。

如图 2 所示, 图 2 为本申请金融数据处理方法第二实施例的流程示意图。

本申请金融数据处理方法第二实施例中, 本实施例在第一实施例的基础上, 所述步骤 S20 之前, 该方法还包括:

S50, 根据预先确定的分类规则, 对各因子数据进行分类处理;

S60, 根据各所述因子数据的种类, 分别建立各所述因子数据对应的因子数据预测模型;

S70, 将建立的各所述因子数据预测模型与其对应的因子数据之间的映射关系进行存储处理。

优选的, 本实施例中, 上述步骤 S60 包括以下步骤:

当一因子数据的种类为第一数据种类时, 则基于以下运算表达式建立所

述因子数据对应的因子数据预测模型:

$$X(K+1) = Z(K)$$

其中, $X(K+1)$ 为 $K+1$ 时刻的因子数据的预测值, $Z(K)$ 为 K 时刻的因子数据的观测值。

上述第一数据种类的因子数据对应的因子数据预测模型的建立方法适用于低频因子数据, 所谓低频因子数据为随价格变化, 缓慢变化的因子数据。

优选的, 本实施例中, 上述步骤 S60 还包括以下步骤:

当一因子数据的种类为第二数据种类时, 所述因子数据对应的因子数据预测模型的建立方法为:

采集第二预设时间区间内的所述因子数据;

对采集的所述因子数据进行预处理(例如, 归一化处理);

利用预处理后的所述因子数据, 且基于长短期记忆神经网络构建所述因子数据对应的因子数据预测模型。

上述基于长短期记忆(LSTM, Long Short-Term Memory)神经网络构建所述因子数据对应的因子数据预测模型的步骤具体为:

基于交叉验证法(cross-validation), 将预处理后的因子数据划分为训练集、评估集和测试集(例如, 70%数量的因子数据作为训练集, 10%数量的因子数据作为评估集, 20%数量的因子数据作为测试集)。将训练集输入至 LSTM 神经网络模型, 且基于梯度下降法(例如, 随机梯度下降法)进行训练。所述评估集用于在训练过程中对 LSTM 神经网络模型进行验证, 将所述测试集输入训练得到的所述 LSTM 神经网络模型, 以对训练得到的所述 LSTM 神经网络模型进行验证, 当训练得到的所述 LSTM 神经网络模型满足第一预设验证条件(例如, 与验证结果差值小于预设阈值), 则训练完成, 将训练完成的 LSTM 神经网络模型设置为该因子数据的因子数据预测模型。

需要注意的是, 上述基于交叉验证法将因子数据划分为训练集、评估集和测试集的步骤可替换为: 基于交叉验证法将因子数据划分为训练集和测试集。且训练集、评估集和测试集中因子数据的数量可根据需要设置, 并不限于上述例举的方案。

上述第二数据种类的因子数据对应的因子数据预测模型的建立方法适用于高频因子数据, 所谓高频因子数据为随价格变化, 迅速变化的因子数据, 这是因为 LSTM 神经网络针对高频因子数据具有更为精准的预测能力。

判断一因子数据属于低频因子数据还是高频因子数据的方法可参照如下示例:

方差法:

取一时间区间内的因子数据, 并对该因子数据以时间为变量作差分处理; 计算差分处理后的因子数据的相对标准差作为变化频率值;

若一因子数据的变化频率值大于或等于预设的标准变化频率值, 则将该因子数据确定为高频因子数据;

若一因子数据的变化频率值小于预设的标准变化频率值，则将该因子数据确定为低频因子数据；

其中，若以股票因子数据为例，所述预设的标准变化频率值可设置为收盘价格的变化频率值。

交叉点法：

取一时间区间内的因子数据，并对该因子数据以时间为变量作差分处理以获得因子数据曲线，该因子数据曲线所在的坐标 X 轴为采样时刻，Y 轴为差分处理后因子数据的值；

在该时间区间内，统计所述因子数据曲线与均值线的交点数量；其中，所述均值为差分处理后因子数据的平均值；

若一因子数据曲线与均值线的交点数量大于或等于预设的标准交点数量，则将该因子数据曲线对应的因子数据确定为高频因子数据；

若一因子数据曲线与均值线的交点数量小于预设的标准交点数量，则将该因子数据曲线对应的因子数据确定为低频因子数据；

其中，若以股票因子数据为例，所述预设的标准交点数量可设置为基于收盘价格差分处理后的收盘价格曲线的标准交点数量。

此外，本实施例中，上述步骤 S50、步骤 S60 及步骤 S70 可由以下步骤替换：

采集第三预设时间区间内的因子数据；

对采集的所述因子数据进行预处理；

将预处理后的所述因子数据按照预设规则划分为训练集及测试集（例如，按照时间先后顺序，划分 80% 数量的因子数据作为训练集，划分 20% 数量的因子数据作为测试集）；

利用训练集，且基于多个预测模型（例如，神经网络模型、随机森林模型、线性回归模型、逻辑回归模型等）进行训练，利用测试集对完成训练的多个预测模型进行验证，根据预先确定的选择规则从上述多个预测模型中选取一预测模型作为该因子数据的因子数据预测模型。

其中，上述预先确定的选择规则可以是选择与验证结果最相近的预测模型作为因子数据预测模型。

上述因子数据预测模型的建立方法适用于各类因子数据（此方案不需对因子数据分类），基于一因子数据同时训练多个预测模型，找出最优的预测模型作为该因子数据的因子数据预测模型。

此外，本申请还提出一种金融数据处理系统。

请参阅图 3，是本申请金融数据处理系统 10 较佳实施例的运行环境示意图。

在本实施例中，金融数据处理系统 10 安装并运行于电子装置 1 中。电子装置 1 可以是桌上型计算机、笔记本、掌上电脑及服务器等计算设备。该电

子装置 1 可包括,但不限于,存储器 11、处理器 12 及显示器 13。图 3 仅示出了具有组件 11-13 的电子装置 1,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。

存储器 11 在一些实施例中可以是电子装置 1 的内部存储单元,例如该电子装置 1 的硬盘或内存。存储器 11 在另一些实施例中也可以是电子装置 1 的外部存储设备,例如电子装置 1 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。进一步地,存储器 11 还可以既包括电子装置 1 的内部存储单元也包括外部存储设备。存储器 11 用于存储安装于电子装置 1 的应用软件及各类数据,例如金融数据处理系统 10 的程序代码等。存储器 11 还可以用于暂时地存储已经输出或者将要输出的数据。

处理器 12 在一些实施例中可以是一中央处理器 (Central Processing Unit, CPU),微处理器或其他数据处理芯片,用于运行存储器 11 中存储的程序代码或处理数据,例如执行金融数据处理系统 10 等。

显示器 13 在一些实施例中可以是 LED 显示器、液晶显示器、触控式液晶显示器以及 OLED (Organic Light-Emitting Diode, 有机发光二极管) 触摸器等。显示器 13 用于显示在电子装置 1 中处理的信息以及用于显示可视化的用户界面。电子装置 1 的部件 11-13 通过系统总线相互通信。

请参阅图 4,是本申请金融数据处理系统 10 较佳实施例的程序模块图。在本实施例中,金融数据处理系统 10 可以被分割成一个或多个模块,一个或多个模块被存储于存储器 11 中,并由一个或多个处理器 (本实施例为处理器 12) 所执行,以完成本申请。例如,在图 4 中,金融数据处理系统 10 可以被分割成第一获取模块 101、查询模块 102、第二获取模块 103 及数据处理模块 104。本申请所称的模块是指能够完成特定功能的一系列计算机程序指令段,比程序更适合于描述金融数据处理系统 10 在电子装置 1 中的执行过程,其中:

第一获取模块 101,用于获取第一预设时间区间内各待处理因子数据的观测值;

查询模块 102,用于在预先建立的因子数据预测模型库中,根据预先确定的各因子数据与因子数据预测模型之间的映射关系,查询各所述待处理因子数据对应的因子数据预测模型;

第二获取模块 103,用于根据各所述待处理因子数据对应的所述因子数据预测模型,获取所述第一预设时间区间内各所述待处理因子数据的预测值;

数据处理模块 104,用于将获取的各所述待处理因子数据的观测值及各所述待处理因子数据的预测值输入预先确定的数据处理模型,以获得修正后的因子数据。

若以股票的因子数据为例,上述因子数据可包括股票当日涨幅、移动平均价 (moving average)、股票交易量等。

用户可根据需要设置第一预设时间区间,例如,设置第一预设时间区间

为 2018 年 1 月 2 号至 2018 年 1 月 22 号。

上述获取第一预设时间区间内各待处理因子数据的观测值的频率可根据需要设置，例如，实时，或者设置固定的时间间隔（例如，1 天），或者在接收到用户发出的获取指令时执行获取第一预设时间区间内各待处理因子数据的观测值的步骤。需要注意的是，一些因子数据（例如，净利润增长、营业利润增长等）若需要通过计算得到，则可预先执行计算步骤，并存储该计算步骤得到的因子数据至预设的存储空间，当执行所述获取第一预设时间区间内各待处理因子数据的观测值的步骤时，第一获取模块 101 在存储空间读取所需的因子数据即可，在一些实施例中，第一获取模块 101 也可执行因子数据的计算步骤。

在查询模块 102 执行器查询步骤之前，金融数据处理系统 10 预先针对各因子数据建立对应的因子数据预测模型，由于各因子数据的特性不同，针对其建立对应的因子数据预测模型的方法也不同，将建立的各因子数据预测模型存储至因子数据预测模型库中，同时保存因子数据与因子数据预测模型之间的映射关系（例如，反映因子数据与因子数据预测模型映射关系的映射表）。

上述第二获取模块 103 获取所述第一预设时间区间内各所述待处理因子数据的预测值的步骤可具体举例为：若要获得 K 时刻的待处理因子数据的预测值，则将 K-1 时刻的待处理因子数据的观测值输入至因子数据预测模型中，因子数据预测模型输出结果即为 K 时刻的待处理因子数据的预测值。

本实施例确定待处理因子数据对应的因子数据预测模型，根据该因子数据预测模型计算出待处理因子数据的预测值，将第一预设时间区间内的因子数据的观测值及预测值输入至数据处理模型，从而得到除噪后的因子数据；相较于现有技术，本实施例充分考虑了各因子数据之间特性不同，针对其采用不同的因子数据预测模型可使因子数据的预测值更为准确、更为符合真实情况，此外，经数据处理模型除噪后的因子数据消除了因子数据观测值中的噪声，提高了因子数据的准确性，为后续针对因子数据的分析奠定了良好基础。

优选的，本实施例中，上述数据处理模块 104 还用于：

将获取的各所述待处理因子数据的观测值按照时间先后顺序进行排序，以生成所述待处理因子数据的观测矩阵；

将获取的各所述待处理因子数据的预测值按照时间先后顺序进行排序，以生成所述待处理因子数据的预测矩阵；

将所述待处理因子数据的观测矩阵及所述待处理因子数据的预测矩阵输入至预先确定的卡尔曼滤波模型，以获得修正后的所述因子数据。

本实施例中，上述生成待处理因子数据的观测矩阵的步骤还可由第一获取模块 101 在执行完所述获取第一预设时间区间内各待处理因子数据的观测值的步骤后立即执行；同样，上述生成待处理因子数据的预测矩阵的步骤还可由第二获取模块 103 在执行完所述获取所述第一预设时间区间内各所述待

处理因子数据的预测值的步骤后立即执行。

本实施例中数据处理模型为卡尔曼滤波模型，卡尔曼滤波模型可高效的消除因子数据的噪声，使因子数据更为接近其真实值。

如图 5 所示，图 5 为本申请金融数据处理系统第二实施例的程序模块图。

本申请金融数据处理系统第二实施例中，本实施例在第一实施例的基础上，所述金融数据处理系统 10 还包括模型构建模块 105，所述模型构建模块 105 用于：

根据预先确定的分类规则，对各因子数据进行分类处理；根据各所述因子数据的种类，分别建立各所述因子数据对应的因子数据预测模型；将建立的各所述因子数据预测模型与其对应的因子数据之间的映射关系进行存储处理。

优选的，本实施例中，上述所述模型构建模块 105 还用于：

当一因子数据的种类为第一数据种类时，则基于以下运算表达式建立所述因子数据对应的因子数据预测模型：

$$X(K+1) = Z(K)$$

其中， $X(K+1)$ 为 $K+1$ 时刻的因子数据的预测值， $Z(K)$ 为 K 时刻的因子数据的观测值。

上述第一数据种类的因子数据对应的因子数据预测模型的建立方法适用于低频因子数据，所谓低频因子数据为随价格变化，缓慢变化的因子数据。

优选的，本实施例中，上述所述模型构建模块 105 还用于：

当一因子数据的种类为第二数据种类时，所述因子数据对应的因子数据预测模型的建立方法为：

采集第二预设时间区间内的所述因子数据；

对采集的所述因子数据进行预处理（例如，归一化处理）；

利用预处理后的所述因子数据，且基于长短期记忆神经网络构建所述因子数据对应的因子数据预测模型。

上述基于长短期记忆（LSTM，Long Short-Term Memory）神经网络构建所述因子数据对应的因子数据预测模型的步骤具体为：

基于交叉验证法(cross-validation)，将预处理后的因子数据划分为训练集、评估集和测试集（例如，70%数量的因子数据作为训练集，10%数量的因子数据作为评估集，20%数量的因子数据作为测试集）。将训练集输入至 LSTM 神经网络模型，且基于梯度下降法（例如，随机梯度下降法）进行训练。所述评估集用于在训练过程中对 LSTM 神经网络模型进行验证，将所述测试集输入训练得到的所述 LSTM 神经网络模型，以对训练得到的所述 LSTM 神经网络模型进行验证，当训练得到的所述 LSTM 神经网络模型满足第一预设验证条件（例如，与验证结果差值小于预设阈值），则训练完成，将训练完成的 LSTM 神经网络模型设置为该因子数据的因子数据预测模型。

需要注意的是，上述基于交叉验证法将因子数据划分为训练集、评估集和测试集的步骤可替换为：基于交叉验证法将因子数据划分为训练集和测试集。且训练集、评估集和测试集中因子数据的数量可根据需要设置，并不限于上述例举的方案。

上述第二数据种类的因子数据对应的因子数据预测模型的建立方法适用于高频因子数据，所谓高频因子数据为随价格变化，迅速变化的因子数据，这是因为 LSTM 神经网络针对时序类的因子数据具有更为精准的预测能力。

此外，本实施例中，上述所述模型构建模块 105 还用于：

采集第三预设时间区间内的因子数据；

对采集的所述因子数据进行预处理；

将预处理后的所述因子数据按照预设规则划分为训练集及测试集（例如，按照时间先后顺序，划分 80% 数量的因子数据作为训练集，划分 20% 数量的因子数据作为测试集）；

利用训练集，且基于多个预测模型（例如，神经网络模型、随机森林模型、线性回归模型、逻辑回归模型等）进行训练，利用测试集对完成训练的多个预测模型进行验证，根据预先确定的选择规则从上述多个预测模型中选取一预测模型作为该因子数据的因子数据预测模型。

其中，上述预先确定的选择规则可以是选择与验证结果最相近的预测模型作为因子数据预测模型。

上述因子数据预测模型的建立方法适用于各类因子数据（此方案不需对因子数据分类），基于一因子数据同时训练多个预测模型，找出最优的预测模型作为该因子数据的因子数据预测模型。

进一步地，本申请还提出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金融数据处理系统，所述金融数据处理系统可被至少一个处理器执行，以使所述至少一个处理器执行上述任一实施例中的金融数据处理方法。

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，凡是在本申请的发明构思下，利用本申请说明书及附图内容所作的等效结构变换，或直接/间接运用在其他相关的技术领域均包括在本申请的专利保护范围内。

权利要求书

1、一种电子装置，其特征在于，所述电子装置包括存储器和处理器，所述存储器上存储有可在所述处理器上运行的金融数据处理系统，所述金融数据处理系统被所述处理器执行时实现如下步骤：

S10，获取第一预设时间区间内各待处理因子数据的观测值；

S20，在预先建立的因子数据预测模型库中，根据预先确定的各因子数据与因子数据预测模型之间的映射关系，查询各所述待处理因子数据对应的因子数据预测模型；

S30，根据各所述待处理因子数据对应的所述因子数据预测模型，获取所述第一预设时间区间内各所述待处理因子数据的预测值；

S40，将获取的各所述待处理因子数据的观测值及各所述待处理因子数据的预测值输入预先确定的数据处理模型，以获得修正后的因子数据。

2、如权利要求1所述的电子装置，其特征在于，在所述步骤S20之前，所述处理器还用于执行所述金融数据处理系统，以实现以下步骤：

S50，根据预先确定的分类规则，对各因子数据进行分类处理；

S60，根据各所述因子数据的种类，分别建立各所述因子数据对应的因子数据预测模型；

S70，将建立的各所述因子数据预测模型与其对应的因子数据之间的映射关系进行存储处理。

3、如权利要求2所述的电子装置，其特征在于，所述步骤S60包括：

当一因子数据的种类为第一数据种类时，则基于以下运算表达式建立所述因子数据对应的因子数据预测模型：

$$X(K+1)=Z(K)$$

其中， $X(K+1)$ 为 $K+1$ 时刻的因子数据的预测值， $Z(K)$ 为 K 时刻的因子数据的观测值。

4、如权利要求2所述的电子装置，其特征在于，所述步骤S60还包括：

当一因子数据的种类为第二数据种类时，所述因子数据对应的因子数据预测模型的建立方法包括：

采集第二预设时间区间内的所述因子数据；

对采集的所述因子数据进行预处理；

利用预处理后的所述因子数据，且基于长短期记忆神经网络构建所述因子数据对应的因子数据预测模型。

5、如权利要求1至4中任意一项所述的电子装置，其特征在于，所述步骤S40包括：

将获取的各所述待处理因子数据的观测值按照时间先后顺序进行排序，以生成所述待处理因子数据的观测矩阵；

将获取的各所述待处理因子数据的预测值按照时间先后顺序进行排序，以生成所述待处理因子数据的预测矩阵；

将所述待处理因子数据的观测矩阵及所述待处理因子数据的预测矩阵输

入至预先确定的卡尔曼滤波模型，以获得修正后的所述因子数据。

6、一种金融数据处理方法，其特征在于，该方法包括步骤：

S10，获取第一预设时间区间内各待处理因子数据的观测值；

S20，在预先建立的因子数据预测模型库中，根据预先确定的各因子数据与因子数据预测模型之间的映射关系，查询各所述待处理因子数据对应的因子数据预测模型；

S30，根据各所述待处理因子数据对应的所述因子数据预测模型，获取所述第一预设时间区间内各所述待处理因子数据的预测值；

S40，将获取的各所述待处理因子数据的观测值及各所述待处理因子数据的预测值输入预先确定的数据处理模型，以获得修正后的因子数据。

7、如权利要求6所述的金融数据处理方法，其特征在于，在所述步骤S20之前，所述金融数据处理方法还包括：

S50，根据预先确定的分类规则，对各因子数据进行分类处理；

S60，根据各所述因子数据的种类，分别建立各所述因子数据对应的因子数据预测模型；

S70，将建立的各所述因子数据预测模型与其对应的因子数据之间的映射关系进行存储处理。

8、如权利要求7所述的金融数据处理方法，其特征在于，所述步骤S60包括：

当一因子数据的种类为第一数据种类时，则基于以下运算表达式建立所述因子数据对应的因子数据预测模型：

$$X(K+1) = Z(K)$$

其中， $X(K+1)$ 为 $K+1$ 时刻的因子数据的预测值， $Z(K)$ 为 K 时刻的因子数据的观测值。

9、如权利要求7所述的金融数据处理方法，其特征在于，所述步骤S60还包括：

当一因子数据的种类为第二数据种类时，所述因子数据对应的因子数据预测模型的建立方法包括：

采集第二预设时间区间内的所述因子数据；

对采集的所述因子数据进行预处理；

利用预处理后的所述因子数据，且基于长短期记忆神经网络构建所述因子数据对应的因子数据预测模型。

10、如权利要求6至9中任一项所述的金融数据处理方法，其特征在于，所述步骤S40包括：

将获取的各所述待处理因子数据的观测值按照时间先后顺序进行排序，以生成所述待处理因子数据的观测矩阵；

将获取的各所述待处理因子数据的预测值按照时间先后顺序进行排序，以生成所述待处理因子数据的预测矩阵；

将所述待处理因子数据的观测矩阵及所述待处理因子数据的预测矩阵输

入至预先确定的卡尔曼滤波模型，以获得修正后的所述因子数据。

11、一种金融数据处理系统，其特征在于，所述金融数据处理系统包括：

第一获取模块，用于获取第一预设时间区间内各待处理因子数据的观测值；

查询模块，用于在预先建立的因子数据预测模型库中，根据预先确定的各因子数据与因子数据预测模型之间的映射关系，查询各所述待处理因子数据对应的因子数据预测模型；

第二获取模块，用于根据各所述待处理因子数据对应的所述因子数据预测模型，获取所述第一预设时间区间内各所述待处理因子数据的预测值；

数据处理模块，用于将获取的各所述待处理因子数据的观测值及各所述待处理因子数据的预测值输入预先确定的数据处理模型，以获得修正后的因子数据。

12、如权利要求 11 所述的金融数据处理系统，其特征在于，所述金融数据处理系统还包括模型构建模块，所述模型构建模块用于：

根据预先确定的分类规则，对各因子数据进行分类处理；

根据各所述因子数据的种类，分别建立各所述因子数据对应的因子数据预测模型；

将建立的各所述因子数据预测模型与其对应的因子数据之间的映射关系进行存储处理。

13、如权利要求 12 所述的金融数据处理系统，其特征在于，所述模型构建模块还用于：

当一因子数据的种类为第一数据种类时，则基于以下运算表达式建立所述因子数据对应的因子数据预测模型：

$$X(K+1) = Z(K)$$

其中， $X(K+1)$ 为 $K+1$ 时刻的因子数据的预测值， $Z(K)$ 为 K 时刻的因子数据的观测值。

14、如权利要求 12 所述的金融数据处理系统，其特征在于，所述模型构建模块还用于：

当一因子数据的种类为第二数据种类时，所述因子数据对应的因子数据预测模型的建立方法包括：

采集第二预设时间区间内的所述因子数据；

对采集的所述因子数据进行预处理；

利用预处理后的所述因子数据，且基于长短期记忆神经网络构建所述因子数据对应的因子数据预测模型。

15、如权利要求 11 至 14 中任意一项所述的金融数据处理系统，其特征在于，所述数据处理模块具体用于：

将获取的各所述待处理因子数据的观测值按照时间先后顺序进行排序，以生成所述待处理因子数据的观测矩阵；

将获取的各所述待处理因子数据的预测值按照时间先后顺序进行排序，

以生成所述待处理因子数据的预测矩阵;

将所述待处理因子数据的观测矩阵及所述待处理因子数据的预测矩阵输入至预先确定的卡尔曼滤波模型, 以获得修正后的所述因子数据。

16、一种计算机可读存储介质, 其特征在于, 所述计算机可读存储介质存储有金融数据处理系统, 所述金融数据处理系统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下步骤:

S10, 获取第一预设时间区间内各待处理因子数据的观测值;

S20, 在预先建立的因子数据预测模型库中, 根据预先确定的各因子数据与因子数据预测模型之间的映射关系, 查询各所述待处理因子数据对应的因子数据预测模型;

S30, 根据各所述待处理因子数据对应的所述因子数据预测模型, 获取所述第一预设时间区间内各所述待处理因子数据的预测值;

S40, 将获取的各所述待处理因子数据的观测值及各所述待处理因子数据的预测值输入预先确定的数据处理模型, 以获得修正后的因子数据。

17、如权利要求 16 所述的计算机可读存储介质, 其特征在于, 在所述步骤 S20 之前, 所述处理器还用于执行所述金融数据处理系统, 以实现以下步骤:

S50, 根据预先确定的分类规则, 对各因子数据进行分类处理;

S60, 根据各所述因子数据的种类, 分别建立各所述因子数据对应的因子数据预测模型;

S70, 将建立的各所述因子数据预测模型与其对应的因子数据之间的映射关系进行存储处理。

18、如权利要求 17 所述的计算机可读存储介质, 其特征在于, 所述步骤 S60 包括:

当一因子数据的种类为第一数据种类时, 则基于以下运算表达式建立所述因子数据对应的因子数据预测模型:

$$X(K+1) = Z(K)$$

其中, $X(K+1)$ 为 $K+1$ 时刻的因子数据的预测值, $Z(K)$ 为 K 时刻的因子数据的观测值。

19、如权利要求 17 所述的计算机可读存储介质, 其特征在于, 所述步骤 S60 还包括:

当一因子数据的种类为第二数据种类时, 所述因子数据对应的因子数据预测模型的建立方法包括:

采集第二预设时间区间内的所述因子数据;

对采集的所述因子数据进行预处理;

利用预处理后的所述因子数据, 且基于长短期记忆神经网络构建所述因子数据对应的因子数据预测模型。

20、如权利要求 16 至 19 中任意一项所述的计算机可读存储介质, 其特

征在于，所述步骤 S40 包括：

将获取的各所述待处理因子数据的观测值按照时间先后顺序进行排序，以生成所述待处理因子数据的观测矩阵；

将获取的各所述待处理因子数据的预测值按照时间先后顺序进行排序，以生成所述待处理因子数据的预测矩阵；

将所述待处理因子数据的观测矩阵及所述待处理因子数据的预测矩阵输入至预先确定的卡尔曼滤波模型，以获得修正后的所述因子数据。

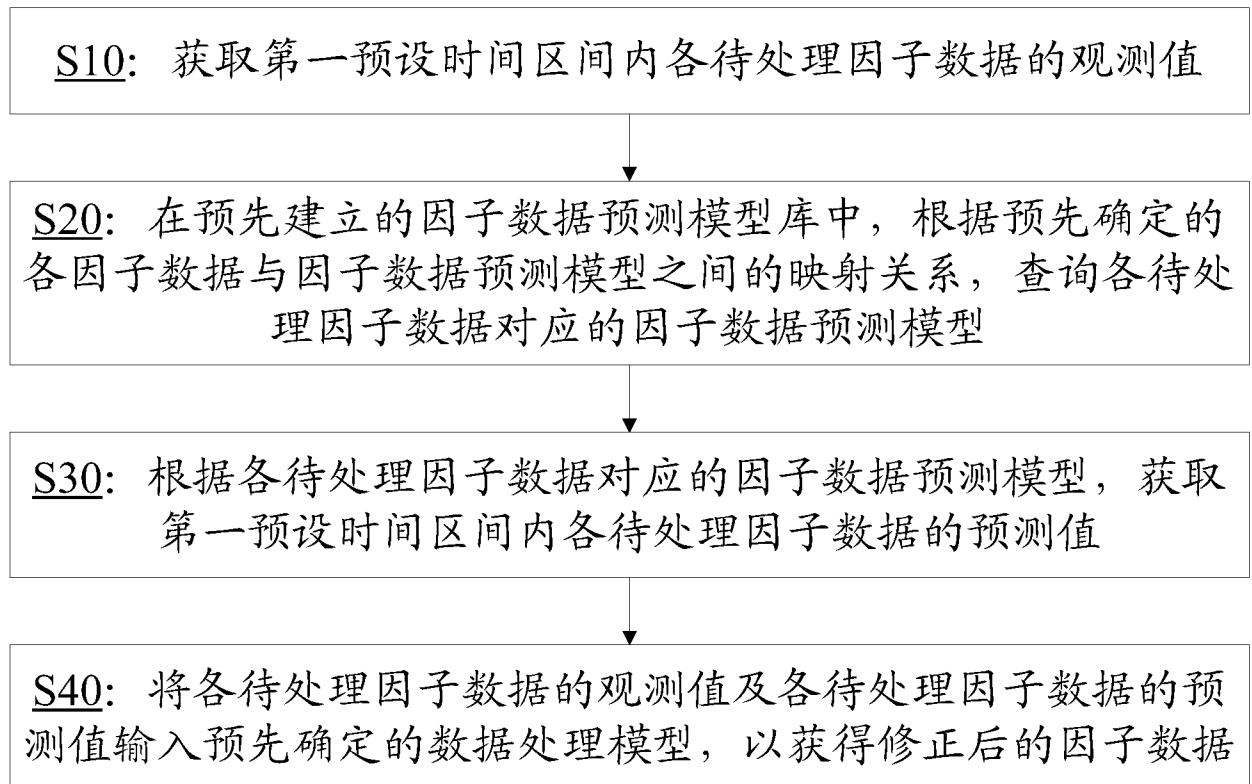


图 1

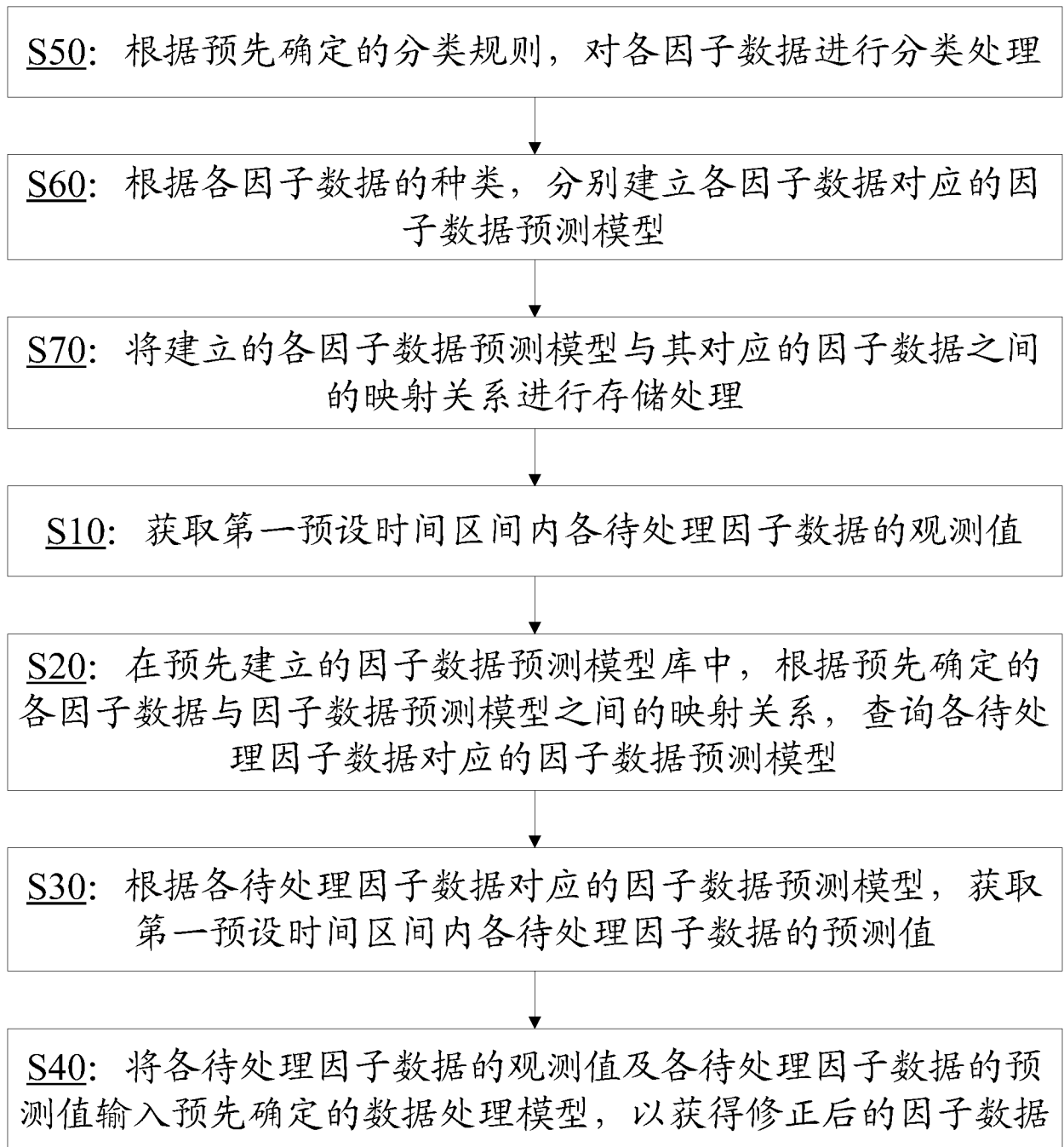


图 2

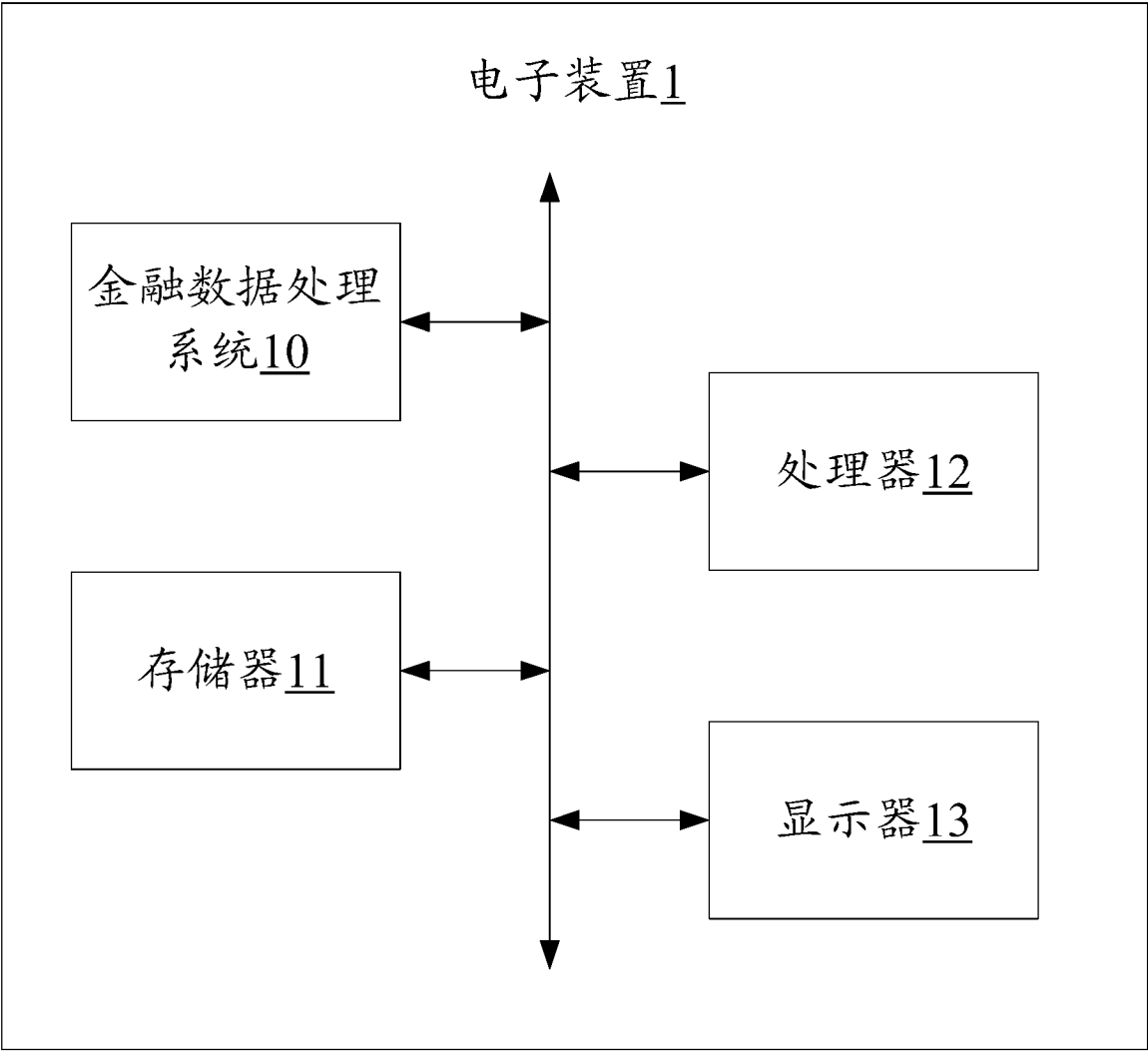


图 3



图 4



图 5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/102226

A. CLASSIFICATION OF SUBJECT MATTER

G06Q 10/00(2012.01)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)

G06Q 10/-

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)

DWPI, CNTXT, USTXT, CPRSABS, SIPOABS, CNKI: 存储器, 数据, 时间, 预测, 模型, 分类, 因子, 金融, memory, storage, data, time, prediction, model, classification, factor, financial,

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 107798604 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 13 March 2018 (2018-03-13) description, paragraphs 27-166	1, 6, 11, 16
Y	CN 107766888 A (ZHONGAN INFORMATION TECHNOLOGY SERVICE CO., LTD.) 06 March 2018 (2018-03-06) description, paragraphs 43-60	1, 6, 11, 16
A	CN 107730087 A (PING AN TECHNOLOGY SHENZHEN CO., LTD.) 23 February 2018 (2018-02-23) entire document	1-20
A	US 2018053102 A1 (TOYOTA JIDOSHA KK) 22 February 2018 (2018-02-22) 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

03 January 2019

Date of mailing of the international search report

14 January 2019

Name and mailing address of the ISA/CN

State Intellectual Property Office of the P. R. China (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/CN2018/102226

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	107798604	A	13 March 2018	None			
CN	107766888	A	06 March 2018	None			
CN	107730087	A	23 February 2018	None			
US	2018053102	A1	22 February 2018	JP	2018027776	A	22 February 2018

国际检索报告

国际申请号

PCT/CN2018/102226

A. 主题的分类

G06Q 10/00(2012.01)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)

G06Q 10/-

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))

DWPI, CNTXT, USTXT, CPRSABS, SIPOABS, CNKI: 存储器, 数据, 时间, 预测, 模型, 分类, 因子, 金融, memory, storage, data, time, prediction, model, classification, factor, financial,

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
Y	CN 107798604 A (平安科技深圳有限公司) 2018年 3月 13日 (2018 - 03 - 13) 说明书第27-166段	1, 6, 11, 16
Y	CN 107766888 A (众安信息技术服务有限公司) 2018年 3月 6日 (2018 - 03 - 06) 说明书第43-60段	1, 6, 11, 16
A	CN 107730087 A (平安科技深圳有限公司) 2018年 2月 23日 (2018 - 02 - 23) 全文	1-20
A	US 2018053102 A1 (TOYOTA JIDOSHA KK) 2018年 2月 22日 (2018 - 02 - 22) 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

2019年 1月 3日

国际检索报告邮寄日期

2019年 1月 14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

受权官员

田竞

电话号码 62089417

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/102226

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	107798604	A	2018年 3月 13日	无			
CN	107766888	A	2018年 3月 6日	无			
CN	107730087	A	2018年 2月 23日	无			
US	2018053102	A1	2018年 2月 22日	JP	2018027776	A	2018年 2月 22日