

## 系统性风险传染机制的动态识别方法、终端及存储介质

### 技术领域

本发明涉及金融系统性风险识别的技术领域，具体涉及系统性风险传染机制的动态识别方法、系统及存储介质。

### 背景技术

系统性风险与我国金融体系安全息息相通，对当代经济稳健高质量发展至关重要，是我国金融风险防控工作中的切要任务。

伴随着当今金融各大市场的高速发展，我国金融机构间的关联形式和渠道呈现出丰富化多样化的特征，由此演变的金融网络的结构也日趋复杂且多变；当某一家金融机构因外部或内部因素爆发危机时，其存在的风险将通过各种联系渠道（如借贷渠道）迅速蔓延至网络中的其他金融机构，严重危害整个金融体系的安全运行。

因此，通过深入研究我国金融市场中的系统性风险，有效识别金融系统性风险传染机制、挖掘其中的系统性重要机构已成为维护金融体系安全的关键问题。

### 发明内容

有鉴于此，有必要针对的问题，提供能够准确度较高的识别出金融市场的风险传染路径并挖掘出系统性重要节点的系统性风险传染机制的动态识别方法、系统及存储介质。

本发明提供了系统性风险传染机制的动态识别方法，包括以下步骤：

S10，获取金融机构结构化数据，并将结构化数据序列化，得到收益率时间序列；以及获取金融机构的非结构化数据，对非结构化数据进行量化，得到投资者情绪变化时间序列；

S20，通过多源数据融合模型对收益率时间序列和投资者情绪变化时间序列进行数据融合，得到多源融合数据时间序列；

S30，基于多源融合数据时间序列，采用 Kruskal 算法动态生成金融市场的系统性风险传染机制，并动态识别系统性重要机构。

可选的，所述步骤 S10 中，获取金融机构的非结构化数据，对非结构化数据进行量化，得到投资者情绪变化时间序列，具体包括：

S101，通过网络爬虫程序获取同一天内多个金融机构的多条个股评论信息；

S102，通过 SVM 情感分类器分别计算每条个股评论信息的情绪分值；其中，计算每条个股评论信息的情绪分值，具体包括：

S1021，对个股评论信息进行预处理，得到投资者情绪相关数据；

S1022，通过中文分词包对投资者情绪相关数据进行分词和去除停用词，得到关键词语组；

S1023，将关键词语组映射到向量空间转换为词向量；

S1024，使用 SVM 情感分类器对向量空间的词向量进行情感分类，得到每条个股评论信息中的情绪分值；所述情感分类包括：积极、消极、中性；

S103，汇总每条个股评论信息的情绪分值，计算出每个金融机构的个股日度投资者情绪指数；

S104，重复执行步骤 S101 至 S103，完成预设天数内的投资者情绪指数时间序列；

S105，计算每个金融机构每个个股的日度投资者情绪变化，获得预设天数内的投资者情绪变化时间序列。

可选的，所述步骤 S1024，使用 SVM 情感分类器对向量空间的词向量进行情感分类，具体包括：

S1024-1，将情感分为积极、消极、中性三类；

S1024-2，对于每一个情感类，分别将其作为+1类，而其余2个情感类的所有样本作为-1类，构造一个 binary SVM 情感分类器；

S1024-3，通过训练集对 binary SVM 情感分类器进行训练，获得文本向量表示的情感指数分别与三类情感的间隔距离，以划分超平面；

S1024-4，将个股评论信息对应的词向量输入到训练好的 binary SVM 情感分类器中，binary SVM 情感分类器根据超平面所在位置输出该个股评论信息分别对应三类情感的概率值，并以此作为个股评论信息中的积极、消极、中性的情绪分值。

可选的，所述步骤 S20，通过多源数据融合模型对收益率时间序列和投资者情绪变化时间序列进行数据融合，得到多源融合数据时间序列；具体包括：

S201，构建基于混合神经网络的多源数据融合模型；所述多源数据融合模型包括：SSA 算法层、DCNN 层、BiLSTM 层、Attention 层和全连接层；

S202，多源数据融合模型对收益率时间序列和投资者情绪变化时间序列进行数据融合，具体包括：

S2021，SSA 算法层通过 SSA 算法对收益率时间序列和投资者情绪变化时间序列进行预处理，获得去除噪声后的标准化时间序列；

S2022，DCNN 层对去除噪声后的标准化时间序列进行卷积运算，提取出时间序列的空间维度特征；

S2023，BiLSTM 层对时间序列数据的空间维度特征进行隐含特征信息的提取，获取时间序列前向、后向中隐含的特征信息，获得时间序列的时间维度特征；

S2024，Attention 层为 BiLSTM 层的隐藏层状态序列分配不同的权重，获得核心关键信息；

S2025，全连接层对关键信息进行学习，筛选对输出序列有用的特征信息，获得最终决策后，得到多源融合数据时间序列。

可选的，所述 S2021，SSA 算法层通过 SSA 算法对收益率时间序列和投资者情绪变化时间序列进行预处理，获得去除噪声后的标准化时间序列；具体包括：

S2021-1，根据窗口长度  $L$  将原始时间序列  $S_n$  进行滞后排列操作，获取其对应的轨迹矩阵  $X$ ，并对轨迹矩阵  $X$  进行奇异值分解；

其中， $S_n$  表示收益率时间序列或投资者情绪变化时间序列；

S2021-2，选取重构信息的奇异值数  $r$ ，重新表述轨迹矩阵  $X$  中的有用信息，并通过对角平均法将重新表述的轨迹矩阵  $X$  转换为标准化后的一维时间序列。

可选的，所述步骤 S2023，BiLSTM 层对时间序列数据的空间维度特征进行隐含特征信息的提取，获取时间序列前向、后向中隐含的特征信息，获得时间序列的时间维度特征；具体包括：

对于任一时刻  $t$ ，BiLSTM 层的输出由前向隐藏层长短期记忆  $M_{fo}$  和后向隐藏层长短期记忆  $M_{ba}$  共同决定。

可选的, 所述步骤 S2024, Attention 层为 BiLSTM 层的隐藏层状态序列分配不同的权重, 获得核心关键信息; 具体包括:

S2024-1, 计算在  $t$  时刻隐藏层状态序列的注意力值;

S2024-2, 利用 Softmax 函数获得归一化注意力权重系数;

S2024-3, 对全部时刻的隐藏层状态进行加权求和运算, 获得注意力机制的输出状态向量为  $Attention(Y_t)$ 。

可选的, 所述步骤 S30, 基于多源融合数据时间序列, 采用 Kruskal 算法动态生成金融市场的系统性风险传染机制, 并动态识别系统性重要机构, 具体包括:

S301, 将多源融合数据时间序列转化为金融机构间的空间距离矩阵  $D$ ;

S302, 采用 Kruskal 算法动态生成金融市场的系统性风险传染机制, 具体包括:

S3021, 基于空间距离矩阵  $D$ , 构建基于多源融合数据相似性的金融网络  $G$ ;

其中, 金融网络  $G$  中, 共包含  $n$  个金融机构; 其网络节点表示金融机构, 空间距离  $d_{ij}$  表示网络节点  $i$  和网络节点  $j$  连边的权重;

S3022, 通过 Kruskal 算法获得一个节点数目为  $n$ , 连边数为  $n - 1$  的子网络 MST; 所述子网络 MST 表示系统性风险传染最可能的传染路径;

S3023, 采用复杂网络技术分析金融市场的系统性风险传染机制的动态变化, 并识别不同时期的系统性重要机构。

本发明还提供了系统性风险传染机制的动态识别系统, 包括:

数据获取模块, 用于获取金融机构结构化数据, 并将结构化数据序列化, 得到收益率时间序列; 以及获取金融机构的非结构化数据, 对非结构化数据进行量化, 得到投资者情绪变化时间序列;

数据融合模块, 用于通过多源数据融合模型对收益率时间序列和投资者情绪变化时间序列进行数据融合, 得到多源融合数据时间序列;

风险识别模块, 用于对数据融合模块输出的多源融合数据时间序列, 采用 Kruskal 算法动态生成金融市场的系统性风险传染机制, 并动态识别系统性重要机构。

本发明还提供了一种计算机可读存储介质，当所述存储介质中的指令由电子设备的处理器执行时，使得电子设备能够执行如上所述的系统性风险传染机制的动态识别方法。

本申请提供的技术方案的优点在于：

本发明与传统的仅使用股票等结构化数据相比，增加了表示投资者微观情感数据的非结构化数据，并将其量化为投资者情绪变化时间序列；同时，使用多源数据融合模型完成结构化数据和非结构化数据的融合，得到多源融合数据时间序列；并采用 Kruskal 算法对多源融合数据时间序列进行运算，动态生成金融市场的系统性风险传染机制，并动态识别系统性重要机构；本发明从更丰富的数据维度揭示金融机构系统性风险的相关因素，更加全面准确的识别出金融市场的风险传染机制，将金融与数据挖掘、自然语言处理、深度学习等计算机技术融合，弥补了我国系统性风险传染机制识别中缺乏微观情感信息等非结构化数据的现有空白，实用性极强。

### 附图说明

图 1 是本发明实施例一提供的系统性风险传染机制的动态识别方法的流程示意图；

图 2 是本发明实施例一提供的系统性风险传染机制的动态识别系统的结构示意图；

图 3 是本发明实施例二提供的系统性风险传染机制的动态识别方法中步骤 S10 的流程示意图；

图 4 为本发明实施例二中 SVM 情感分类器的基本分类原理图；

图 5 是本发明实施例三提供的系统性风险传染机制的动态识别方法中步骤 S20 的流程示意图；

图 6 为本发明实施例三中多源数据融合模型的结构示意图；

图 7 为本发明实施例三中引入注意力机制后的 BiLSTM 层的结构示意图；

图 8 是本发明实施例四提供的系统性风险传染机制的动态识别方法中步骤 S30 的流程示意图；

图中：

10 为数据获取模块，20 为数据融合模块，30 为风险识别模块。

### 具体实施方式

本申请中，针对识别金融市场的系统性风险传染机制这一问题，现有研究中所使用的数据大多数为资产负债数据、股票数据等结构化数据，忽略了非结构化数据中包含的微观行为信息；这类微观行为数据，如：投资者根据与某家金融机构相关信息的主观判断而出现的较为普遍的悲观情绪，容易招致金融市场中出现大规模的挤兑现象，进而通过间接传染渠道引发系统性的金融风险危机。

因此，对于系统性风险分析来讲，投资者的情绪信息等非结构化数据能够挖掘出如投资者情绪变化等关键的风险影响因素，是系统性风险分析中至关重要的信息来源。

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的一部分实施例，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作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。

#### 实施例一

参阅图 1，系统性风险传染机制的动态识别方法，包括以下步骤：

S10，获取金融机构结构化数据，并将结构化数据序列化，得到收益率时间序列；以及获取金融机构的非结构化数据，对非结构化数据进行量化，得到投资者情绪变化时间序列；

S20，通过多源数据融合模型对收益率时间序列和投资者情绪变化时间序列进行数据融合，得到多源融合数据时间序列；

S30，基于多源融合数据时间序列，采用 Kruskal 算法动态生成金融市场的系统性风险传染机制，并动态识别系统性重要机构。

需要说明的是，本申请步骤 S10 中，金融机构的结构化数据可从 wind 数据库中获取，所述的结构化数据日度市场交易数据，即个股日度收盘价，考虑到个股收益率数据的平滑性，本发明采用对数收益率计算方式进一步计算个股日

度收益率；个股日度收益率的表达式为：

$$R_t = \ln \left( P_t / P_{t-1} \right);$$

上式中， $R_t$ 表示第 $t$ 个日度的对数收益率， $P_t$ 表示 $t$ 个日度的收盘价， $P_{t-1}$ 表示第 $t-1$ 个日度期的收盘价。

此外，本发明还提供了系统性风险传染机制的动态识别系统。

参阅图 2，系统性风险传染机制的动态识别系统，包括：

数据获取模块 10，用于获取金融机构结构化数据，并将结构化数据序列化，得到收益率时间序列；以及获取金融机构的非结构化数据，对非结构化数据进行量化，得到投资者情绪变化时间序列；

数据融合模块 20，用于通过多源数据融合模型对收益率时间序列和投资者情绪变化时间序列进行数据融合，得到多源融合数据时间序列；

风险识别模块 30，用于对数据融合模块 20 输出的多源融合数据时间序列，采用 Kruskal 算法动态生成金融市场的系统性风险传染机制，并动态识别系统性重要机构。

本发明实施例一提供的系统性风险传染机制的动态识别方法、系统中，与传统的仅使用股票等结构化数据相比，增加了表示投资者微观情感数据的非结构化数据，并将其量化为投资者情绪变化时间序列；同时，使用多源数据融合模型完成结构化数据和非结构化数据的融合，得到多源融合数据时间序列；并采用 Kruskal 算法对多源融合数据时间序列进行运算，动态生成金融市场的系统性风险传染机制，并动态识别系统性重要机构；本发明从更丰富的数据维度揭示金融机构系统性风险的相关因素，更加全面准确的识别出金融市场的风险传染机制，将金融与数据挖掘、自然语言处理、深度学习等计算机技术融合，弥补了我国系统性风险传染机制识别中缺乏微观情感信息等非结构化数据的现有空白。

以上是对本发明提供的系统性风险传染机制的动态识别方法、系统的一个实施例进行的说明，以下将对本发明提供的系统性风险传染机制的动态识别方法的另一个实施例进行说明。

## 实施例二

参阅图 3，在实施例一的基础上，系统性风险传染机制的动态识别方法，所述步骤 S10，获取金融机构的非结构化数据，对非结构化数据进行量化，得到投资者情绪变化时间序列，具体包括：

S101，通过网络爬虫程序获取同一天内多个金融机构的多条个股评论信息。

需要说明的是，本申请步骤 S101 中，金融机构的非结构化数据可通过网络爬虫程序获取同一天内多个金融机构的多条个股评论信息，如：从个股股票论坛获取相关信息进行量化分析研究，获取相关投资者对个股发表评论的舆论数据，此类评论数据可以直接反应出投资者对相应个股行情的情绪态度；值得注意的是，如果投资者在某时期对几家上市金融机构的情绪态度类似，说明这两家上市金融机构在极大可能下具有高度的相关性；

每条帖子（个股评论信息）的抓取字段包括：帖子标题、阅读量、评论量、具体内容、作者名称、发帖时间以及最后更新时间。

对多条个股评论信息进行获取后，利用 SVM 情感分类器量化这些帖子中的投资者情绪，最终获得投资者情绪指数时间序列。

S102，通过 SVM 情感分类器分别计算每条个股评论信息的情绪分值；其中，计算每条个股评论信息的情绪分值，具体包括：

S1021，对个股评论信息进行预处理，得到投资者情绪相关数据。

S1022，通过中文分词包对投资者情绪相关数据进行分词和去除停用词，得到关键词语组。

需要说明的是，本申请步骤 S1022 中，通过中文分词包对投资者情绪相关数据进行分词和去除停用词，得到关键词语组中，可采用中文分词包（Jieba）对情绪帖子进行分词处理以及去停用词操作，分词与去除停用词的处理结果如表 1 所示。

表 1 分词、去除停用词结果示例

原句	这个股为什么反其道而行之。已经下跌一个月了。按说也应该走的差不多了。在最近几天大盘大跌的情况下他怎么还稳定了价格。在大盘大涨的时候他反而大量出货了。这个股有妖气。很害怕。
----	---



## 说明书

分词处理	这个 股 为什么 反其道而行之 。 已经 下跌 一个月 了 。 按说 也 应该 走 的 差不多了 。 在 最近 几天 大盘 大跌 的 情况 下 他 怎么 还 稳定 了 价格 。 在 大盘 大涨 的 时候 他 反而 大量 出 货 了 。 这个 股有 妖气 。 很 害怕 。
去停用词处理	股 反其道而行之 下跌 一个月 走 几天 大盘 大跌 情况 稳定 价格 大盘 大涨 大量 出货 股有 妖气 害怕

S1023, 将关键词语组映射到向量空间转换为词向量。

需要说明的是, 本申请步骤 S1023 中, 将关键词语组映射到向量空间转换为词向量中, 可使用 Python 第三方库 Gensim 中的 Word2Vec 方法对非结构化文本进行处理; Word2Vec 是以无监督方式从海量文本语料中学习富含语义信息的低维词向量的语言模型, Word2Vec 词向量可以用于词语之间相似性度量, 由于语义相近的词语在向量空间上的分布比较接近, 可以通过计算词向量间的空间距离来表示词语间的语义相似度, 因此, Word2Vec 词向量具有很好的语义特性。

S1024, 使用 SVM 情感分类器对向量空间的词向量进行情感分类, 得到每条个股评论信息中的情绪分值; 所述情感分类包括: 积极、消极、中性; 具体包括:

S1024-1, 将情感分为积极、消极、中性三类;

S1024-2, 对于每一个情感类, 分别将其作为+1类, 而其余2个情感类的所有样本作为-1类, 构造一个 binary SVM 情感分类器;

S1024-3, 通过训练集对 binary SVM 情感分类器进行训练, 获得文本向量表示的情感指数分别与三类情感的间隔距离, 以划分超平面;

所述训练集的输入序列的表达式为:

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i), \dots, (x_n, y_n)\}; \quad (1.5)$$

式 (1.5) 中, n 表示文本总数量,  $x_i$  表示第*i*个文本的向量表示,  $y_i$  表示第*i*个文本的情感指数;

S1024-4, 将个股评论信息对应的词向量输入到训练好的 binary SVM 情感分类器中, binary SVM 情感分类器根据超平面所在位置输出该个股评论信息分别对应三类情感的概率值, 并以此作为个股评论信息中的积极、消极、中性的情绪分值。

需要说明的是, 本申请步骤 S1024 中, 使用 SVM 情感分类器对向量空间的词向量进行情感分类中, 使用的 SVM 情感分类器是一种二分类模型, 它的基本模型是定义在特征空间上的间隔最大的线性分类器。

请参阅图 4, SVM 的基本原理为: 在空间中寻找一个超平面, 该超平面能够将空间中的点以最大间隔分为不同的类别, 该超平面满足以下特点:

- (1) 某一类别所有点到达超平面的距离之和为最小值;
- (2) 距离超平面最近的点与超平面的距离为最大值;

其中, 在超空间中任意一点  $x$  到超平面的距离  $r$  为:

$$r = \frac{|w^T x + b|}{\|w\|};$$

上式中,  $w$  表示超平面斜率,  $b$  表示超平面截距;  $w * x + b = 0$  表示线性分割超平面。

本申请中, 将文本的情感分成积极、消极、中性三类, 在处理时, 选择 one-against-all (一对多) 的方法, 对于每一个类, 分别将其作为 +1 类, 而其余 2 个类的所有样本作为 -1 类, 构造一个 binary SVM; 具体为:

将积极作为 +1 类, 将消极与中性表示为 -1 类; 将中性作为 +1 类, 将消极与积极表示为 -1 类; 将消极作为 +1 类, 将积极与中性表示为 -1 类。

通过训练集对 binary SVM 情感分类器进行训练, 使 binary SVM 情感分类器计算文本向量表示的情感指数分别与三类情感的间隔距离, 从而划分超平面。

在对个股评论信息进行情感分类时, 训练好的 binary SVM 情感分类器根据超平面所在位置, 输出个股评论信息分别对应三类情感的概率值, 并以此作为其积极、消极、中性的情绪分值。

S103, 汇总每条个股评论信息的情绪分值, 计算出每个金融机构的个股日度投资者情绪指数; 所述个股日度投资者情绪指数的表达式为:

$$Investor_{i,t} = \sum_{i=1}^N \frac{positive_{i,t} + negative_{i,t} + neutral_{i,t}}{n_{i,t}}; \quad (1.1)$$

式 (1.1) 中,  $Investor_{i,t}$  表示个股  $i$  第  $t$  个日度的投资者情绪指数;

$positive_{i,t}$ 、 $negative_{i,t}$  和  $neutral_{i,t}$  分别表示第  $t$  个日度的个股  $i$  中积极情绪分值、消极情绪分值和中立情绪分值,  $n_{i,t}$  表示第  $t$  个日度的个股  $i$  评论信息中特征词的数量,  $N$  代表个股  $i$  在第  $t$  个日度的全部评论信息;

S104, 重复执行步骤 S101 至 S103, 完成预设天数内的投资者情绪指数时间序列; 所述投资者情绪指数时间序列的表达式为:

$$Inv_i = (Investor_{i,1}, Investor_{i,j}, \dots, Investor_{i,end}); \quad (1.2)$$

式 (1.2) 中,  $Investor_{i,j}$  表示个股  $i$  在预设天数内第  $j$  个日度的情绪指数;

S105, 计算每个金融机构每个个股的日度投资者情绪变化, 获得预设天数内的投资者情绪变化时间序列; 所述投资者情绪变化时间序列的表达式为:

$$I_i = (I_{i,1}, I_{i,2}, \dots, I_{i,t}, \dots, I_{i,end}); \quad (1.3)$$

式 (1.3) 中,  $I_{i,t}$  表示个股  $i$  第  $t$  个日度与第  $t-1$  个日度相比的情绪变化, 其表达式为:

$$I_{i,t} = \ln \left( \frac{Investor_t}{Investor_{t-1}} \right); \quad (1.4)。$$

需要说明的是, 本实施例中, 预设天数可为 1 个月。

本发明实施例二提供的系统性风险传染机制的动态识别方法中, 采用网络爬虫技术对中金融市场论坛 (如: 东方财富网) 的个股评论信息进行抓取, 并采用 SVM 情感分类器分析并量化个股评论信息中的投资者情绪并获得投资者情绪变化时间序列, SVM 情感分类器具有分类结果准确度高的优点, 能够后续多源数据融合提供准确度较高的数据, 进而提高整个系统性风险传染机制识别的准确度。

以上是对本发明提供的系统性风险传染机制的动态识别方法的一个实施例进行的说明, 以下将对本发明提供的系统性风险传染机制的动态识别方法的另一个实施例进行说明。

### 实施例三

参阅图 5，在实施例一的基础上，系统性风险传染机制的动态识别方法，所述步骤 S20，通过多源数据融合模型对收益率时间序列和投资者情绪变化时间序列进行数据融合，得到多源融合数据时间序列；具体包括：

S201，构建基于混合神经网络的多源数据融合模型；所述多源数据融合模型包括：SSA 算法层、DCNN 层、BiLSTM 层、Attention 层和全连接层。

S202，多源数据融合模型对收益率时间序列和投资者情绪变化时间序列进行数据融合，具体包括：

S2021，SSA 算法层通过 SSA 算法对收益率时间序列和投资者情绪变化时间序列进行预处理，获得去除噪声后的标准化时间序列；具体包括：

S2021-1，根据窗口长度  $L$  将原始时间序列  $S_n$  进行滞后排列操作，获取其对应的轨迹矩阵  $X$ ，并对轨迹矩阵  $X$  进行奇异值分解；

其中， $S_n$  表示收益率时间序列或投资者情绪变化时间序列，其表达式为：

$$S_n = (y_1, y_2, \dots, y_n); \quad (2.1)$$

所述轨迹矩阵  $X$  的阶数为  $L \times K$ ，其表达式为：

$$X = \begin{bmatrix} y_1 & y_2 & \cdots & y_k \\ y_2 & y_3 & \cdots & y_{k+1} \\ \vdots & \vdots & & \vdots \\ y_L & y_{L+1} & \cdots & y_n \end{bmatrix}; \quad (2.2)$$

式 (2.2) 中， $K = n - L + 1$ ， $n$  表示原始时间序列  $S_n$  的长度；

对轨迹矩阵  $X$  进行奇异值分解的表达式为：

$$X = U_{L \times L} \begin{pmatrix} \Delta & 0 \\ 0 & 0 \end{pmatrix} V_{K \times K}^T; \quad (2.3)$$

式 (2.3) 中， $\Delta = \text{diag}(\beta_1, \beta_2, \dots, \beta_d)$ ；

$\begin{pmatrix} \Delta & 0 \\ 0 & 0 \end{pmatrix}$  表示  $L \times K$  的对角矩阵， $\Delta$  表示对角矩阵中对角线上的所有元素； $\beta_d$  表

示奇异值，且  $\beta_d > 0$ ；其中， $\beta_1, \beta_2, \dots, \beta_d$  为降序排列；

$U$  表示轨迹矩阵  $X$  的左奇异矩阵，其阶数为  $L \times L$ ； $V$  表示轨迹矩阵  $X$  的阶数右奇异矩阵，其阶数为  $K \times K$ ；

S2021-2, 选取重构信息的奇异值数 $r$ , 重新表述轨迹矩阵 $X$ 中的有用信息, 并通过对角平均法将重新表述的轨迹矩阵 $X$ 转换为标准化后的一维时间序列;

重新表述的轨迹矩阵 $X$ 的表达式为:

$$X \approx U_{L \times r} \Delta_{r \times r} V_{r \times K}^T = (h_{i,j})_{L \times K}; \quad (2.4)$$

式(2.4)中,  $r$ 表示重构信息的奇异值数,  $U_{L \times r}$ 表示前 $r$ 个能够反映目标有用信息的左奇异矩阵,  $V_{r \times K}$ 表示前 $r$ 个能够反映目标有用信息的右奇异矩阵,  $h_{i,j}$ 表示近似轨迹矩阵 $X$ 的元素值;  $\Delta_{r \times r}$ 表示前 $r$ 个能够反映目标有用信息的奇异值元素;

标准化后的一维时间序列的表达式为:

$$y_{rck} = (y_{rc1}, y_{rc2}, \dots, y_{rcn}); \quad (2.5)$$

式(2.5)中,  $y_{rck}$ 的计算表达式为:

$$y_{rck} = \begin{cases} \frac{1}{k} \sum_{m=1}^k h_{m,k-m+1}, 1 \leq k \leq L_p \\ \frac{1}{L_p} \sum_{m=1}^{L_p} h_{m,k-m+1}, L_p \leq k \leq K_p \\ \frac{1}{T-k+1} \sum_{m=k-K_p+1}^{T-K_p+1} h_{m,k-m+1}, K_p \leq k \leq T \end{cases}; \quad (2.6)$$

式(2.6)中,  $L_p = \min(L, K)$ ,  $K_p = \max(L, K)$ 。

S2022, DCNN 层对去除噪声后的标准化时间序列进行卷积运算, 提取出时间序列的空间维度特征。

需要说明的是, 本申请 S2022 中, DCNN 层包括卷积层、批归一化层以及池化层, DCNN 层的具体功能如下: 卷积层通过卷积操作能够捕获数据中抽象的空间特征向量; 批归一化层一方面通过将上述得到的特征向量归一化处理提高数据分布的稳定性, 另一方面能够提升训练模型速度; 为了实现上一层即批一归化层的特征降维同时保证主体信息特征, 本申请使用最大池化(max pooling)的方式完成池化层的处理。

S2023, BiLSTM 层对时间序列数据的空间维度特征进行隐含特征信息的提取, 获取时间序列前向、后向中隐含的特征信息, 获得时间序列的时间维度特征;

需要说明的是, 本申请步骤 S2023 中, 鉴于 CNN 层无法获取时间序列的长

期特征信息，因此本申请中，引入 BiLSTM 结构；本实施例中的 BiLSTM 层由前向和后向的 LSTM 层连接构成。

对于任一时刻  $t$ ，BiLSTM 层的输出由 LSTM 层的两个方向相反的传播状态共同决定，正向传播能够从前向后捕捉“过去”时刻的信息，得到前向隐藏层长短期记忆  $M_{fo}$ ；反向传播能够从前向后捕捉“未来时刻”的信息，得到后向隐藏层长短期记忆  $M_{ba}$ ；即：

$$Y_t = M_{fo} \oplus M_{ba}; \quad (2.7)$$

式 (2.7) 中， $Y_t$  表示  $t$  时刻的时间序列的时间维度特征。

考虑到不同时刻 BiLSTM 的输入有不同的重要程度，本申请引入注意力机制，引入注意力机制后的 BiLSTM 层的结构示意图，请参阅图 7。

S2024，Attention 层为 BiLSTM 层的隐藏层状态序列分配不同的权重，获得核心关键信息；具体包括：

S2024-1，计算在  $t$  时刻隐藏层状态序列的注意力值  $l_t^i$ ，计算表达式为：

$$l_t^i = M_a^T \tanh(W_a c_{i-1} + U_a Y_t); \quad (2.8)$$

式 (2.8) 中， $Y_t$  表示 BiLSTM 层的隐藏层输出的特征向量序列；

$M_a$ 、 $W_a$ 、 $U_a$  表示输入的权重矩阵； $c_{i-1}$  为解码器  $t-1$  时刻的状态；

S2024-2，利用 Softmax 函数获得归一化注意力权重系数  $\alpha_t^i$ ；

计算表达式为：

$$\alpha_t^i = \frac{\exp(l_t^i)}{\sum_{i=1}^T \exp(l_t^i)}; \quad (2.9)$$

式 (2.9) 中， $\alpha_t^i$  表示归一化注意力权重系数；

S2024-3，对全部时刻的隐藏层状态  $Y_t$  进行加权求和运算，获得注意力机制的输出状态向量为  $Attention(Y_t)$ ；

$Attention(Y_t)$  的计算表达式为：

$$Attention(Y_t) = \sum_{i=1}^T \alpha_t^i Y_t^i; \quad (2.10)。$$

需要说明的是，本申请步骤 S2024 中，注意力机制能够通过计算不同时刻隐藏层状态的注意力值、借助归一化处理，计算对应的注意力权重系数以及对不同时刻隐藏状态序列进行加权求和，实现集中精力学习更重要更关键的特征

目标。

S2025, 全连接层对关键信息进行学习, 筛选对输出序列有用的特征信息, 获得最终决策后, 得到多源融合数据时间序列。

本发明实施例一提供的系统性风险传染机制的动态识别方法中, 由于收益率时间序列以及投资者情绪变化时间序列的信息维度大且噪声多, 需要通过 SSA 算法进行序列去噪处理, 然后将进行去噪处理后的时间序列作为 DCNN 层的输入, DCNN 层通过其卷积层的卷积运算提取时间序列的空间维度特征后, 将其输入到 BiLSTM 层, BiLSTM 层通过学习正向传播状态的 LSTM 以及反向传播状态的 LSTM, 获取时间序列前向、后向中隐含的特征信息, 提取出时间序列的时间维度特征信息; 随后, 获得时间序列的空间维度特征; Attention 层通过引入注意力机制, 对 BiLSTM 隐藏层状态序列分配不同的权重, 进一步筛选核心关键信息; 全连接层将综合上述特征的关键信息获得最终决策。

以上是对本发明提供的系统性风险传染机制的动态识别方法、系统的一个实施例进行的说明, 以下将对本发明提供的系统性风险传染机制的动态识别方法的另一个实施例进行说明。

#### 实施例四

参阅图 8, 在实施例一的基础上, 系统性风险传染机制的动态识别方法, 所述步骤 S30, 基于多源融合数据时间序列, 采用 Kruskal 算法动态生成金融市场的系统性风险传染机制, 并动态识别系统性重要机构, 具体包括:

S301, 将多源融合数据时间序列转化为金融机构间的空间距离矩阵  $D$ ;

空间距离矩阵  $D$  中, 元素  $d_{ij}$  的表达式为:

$$d_{ij} = \sqrt{2(1 - \rho_{ij})}; \quad (3.1)$$

式 (3.1) 中,  $i$  和  $j$  分别代表不同的金融机构,  $d_{ij}$  表示金融机构  $i$  和  $j$  的空间距离, 满足  $d_{ij} \in [0, 1]$ ;

$\rho_{ij}$  度量了金融机构  $i$  和  $j$  的基于多源融合数据时间序列的相关性, 其表达式为:

$$\rho_{ij} = \frac{E(Y_i Y_j) - E(Y_i)E(Y_j)}{\sqrt{[E(Y_i^2) - E(Y_i)^2][E(Y_j^2) - E(Y_j)^2]}}; \quad (3.2)$$

式 (3.2) 中,  $E(\cdot)$  为多源融合时间序列的均值;

$E(Y_i Y_j) - E(Y_i)E(Y_j)$  为金融机构  $i$  和  $j$  的协方差;

$\sqrt{[E(Y_i^2) - E(Y_i)^2][E(Y_j^2) - E(Y_j)^2]}$  为金融机构  $i$  和  $j$  标准差的乘积,  $\rho_{ij}$  满足  $\rho_{ij} \in [-1, 1]$ ;

S302, 采用 Kruskal 算法动态生成金融市场的系统性风险传染机制, 具体包括:

S3021, 基于空间距离矩阵  $D$ , 构建基于多源融合数据相似性的金融网络  $G$ ;

其中, 金融网络  $G$  中, 共包含  $n$  个金融机构; 其网络节点表示金融机构, 空间距离  $d_{ij}$  表示网络节点  $i$  和网络节点  $j$  连边的权重;

S3022, 通过 Kruskal 算法获得一个节点数目为  $n$ , 连边数为  $n - 1$  的子网络 MST, 获得金融市场的系统性风险传染机制; 其中, 所述子网络 MST 表示系统性风险传染最可能的传染路径。

需要说明的是, 本申请步骤 S30221 中, 通过 Kruskal 算法处理金融网络  $G$ , 获得一个节点数目为  $n$ , 连边数为  $n-1$  的子网络 MST; 所述子网络 MST 揭示了基于多源融合数据相似性的金融网络中最紧密最重要的连接; 该子网络 MST 意味着在金融市场网络中系统性风险的最有可能的传染机制, 即系统性风险将沿着 MST 中的路径迅速在金融市场中传染扩散。

S3023, 采用复杂网络技术分析金融市场的系统性风险传染机制的动态变化, 并识别不同时期的系统性重要机构; 具体包括:

通过节点重要性维度和传染机制整体结构维度对步骤 S3022 生成的系统性风险传染机制进行拓扑分析, 识别出不同时期的系统性重要节点。

需要说明的是, 对于节点重要性维度, 采用度中心性以及中介中心性来识别传染机制中的系统性重要机构; 其中, 度中心性能够识别出与其他金融机构联系最多的金融机构, 中介中心性能够捕捉到对传染机制中控制程度较高的金融机构;

对于传染机制整体结构维度, 采用归一化树长来衡量系统性风险传染机制的动态长度变化, 即连接紧密性的动态变化。

本发明实施例三提供的系统性风险传染机制的动态识别方法中, 基于多源



融合数据的时间序列相关性以及 MST 方法的网络构建，一方面能够解决由于大多数金融机构间的交易数据不公导致部分数据无法获得的问题，以及传统基于相关性网络构建中仅考虑结构化数据而忽略非结构化数据中包含的微观情感信息的问题；另一方面能够提取出隐含在金融市场中最关键最紧密的联系渠道，即识别出系统性风险最有可能的传染机制。

此外，本发明还提供了一种计算机可读存储介质，当所述存储介质中的指令由电子设备的处理器执行时，使得电子设备能够执行如上所述的系统性风险传染机制的动态识别方法。

所述计算机可读存储介质，可包括：U 盘、移动硬盘、只读存储器 (ROM, Read-OnlyMemory)、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。

综上，本申请中，通过深入分析系统性风险传染机制，可以从宏观上揭示和发现系统性风险在金融网络内的传染机制的动态变化，也可以从微观上有效识别并重点监控系统重要性机构；与此同时，通过重点关注关键时期 (如 2015 年我国股灾、2020 年新冠疫情等时期)，有助于挖掘系统性风险传染机制在关键时期的变化规律，一方面能够为金融机构自身的风险防范提供相应警示，另一方面能够为投资者以及监管当局提供针对性建议。

本申请在识别金融市场的系统性风险传染机制过程中，针对现有研究仅使用结构化数据而忽略了非结构化数据中包含的微观情感信息；提出了一种基于结构化数据和非结构化数据的系统性风险传染机制的动态识别方法，从更丰富的数据维度揭示金融机构系统性风险的相关因素，更加全面准确的识别出金融市场的风险传染机制，弥补了我国系统性风险传染机制识别中缺乏微观情感信息等非结构化数据的现有空白，实用性极强。

以上所述实施例仅表达了本发明的几种实施方式，其描述较为具体和详细，但并不能因此而理解为对本发明专利范围的限制。应当指出的是，对于本领域的普通技术人员来说，在不脱离本发明构思的前提下，还可以做出若干变形和改进，这些都属于本发明的保护范围。因此，本发明的保护范围应以所附权利要求为准。