Market Abnormality Period Detection via Co-movement Attention Model

Conference Paper · December 2018

DOI: 10.1109/BigData.2018.8621877

Market Abnormality Period Detection via Co-movement Attention Model

Yue Wang*, Chenwei Zhang†, Shen Wang†, Philip S Yu†, Lu Bai*, and Lixin Cui*
*Department of Computer Science, Central University of Finance and Economics
Email: wangyuecs, bailucs@cufe.edu.cn, cuilixindora@yahoo.com
†Department of Computer Science, University of Illinois at Chicago
Email: czhang99, swang224, psyu@uic.edu

***Abstract*—The financial contagion describes a widespread phenomenon of the interdependency for pairs of stock time series during the market abnormality periods. Since the interdependency rule between stocks varies in different periods, it is difficult to capture the interdependency rule for stocks related to the market status effectively. We define this interdependency rule as, the co-movement pattern, a high-dimensional co-varying relationship between pairs of stock time series and propose a Co-movement Attention Model (CAM) to discover the co-movement patterns for the stocks related to the market status. With the discovered patterns, CAM focuses on the stock-level abnormality periods by the co-movement attention automatically. CAM is trained under the supervision of the stock sector label information. CAM has the ability to model financial contagion and detect global market abnormality periods, by modeling co-movement patterns on every pair-wise stocks. We verify our methods on the real-world stock data and compare it with state of the art methods. The experimental result shows that our method not only captures the co-movement attentions with better quantitative metric values but also covers more real market abnormalities than the other alternatives.**

***Index Terms*—Financial data mining, co-movement pattern, co-movement attention, stock-level abnormality, market-level abnormality**

I. INTRODUCTION

The financial contagion [1] describes the phenomenon when the interdependency on the changing trends of pairs of stocks becomes strong and widespread in a market or even across different markets during the market abnormality periods (e.g. market crashes or bear market). This phenomenon relates to the interdependency patterns and the co-occurring temporal contextual information for all the stock time series in a market. However, since most existing works [2] [3] neither fully address the interdependency patterns nor utilize the temporal contextual information of stocks, they may face the following challenges when studying the financial contagion in the real-world market.

- **Effective interdependency pattern learning.** Many works apply the pre-defined framework (e.g. correlation [4], wavelet scoring [5], etc.) to discuss the interdependency pattern between a pair of stock time series. However, since these methods discard the information beyond their pre-defined framework, it is very challenging to figure out whether or not these methods capture the interdependency patterns between stock time series effectively.
- **Automatically stock-level abnormality periods detecting.** With the obtained interdependency patterns for a pair of stock time series, recent state of the art methods apply a sliding window framework [2] [3] to detect the stock-level abnormality periods (or time windows) with the high interdependency degrees for a specific pair of stocks. However, since the sliding window framework ignores the temporal contextual information along the time dimension of stock time series between the time windows, it is difficult to automatically detect the stock-level abnormality periods along the time dimension with these methods.
- **Automatically market-level abnormality periods detecting.** Although the abnormality or outlier detection in time series [6] [7] [8] addresses a hot issue in data mining field, none of them aims to detect the market abnormality periods based on the interdependency patterns between pairs of stocks. According to the rules about the financial contagion, it is promising to detect the market-level abnormality by fusing the stock-level abnormality periods, and in turn, the consistency of the detected abnormalities and real market abnormality events can also be used to measure the quality of the obtained stock interdependency patterns. However, since most existing works do not relate the stock-level abnormalities with the market-level abnormalities, it is also a challenge to design a mechanism to utilize all the stock-level abnormalities to automatically reveal the undistorted market-level abnormalities which are consistent with the real market abnormality events.

To address the aforementioned challenges, we formalize the interdependency rule between a pair of stock time series as, the co-movement pattern, a corresponding high-dimensional co-varying relationship. Then, we propose a Co-movement Attention Model, namely CAM, to recognize the co-movement patterns and focus on the stock-level abnormality periods for the stocks in a market automatically. CAM captures the co-movement patterns and collects the evidence satisfied the co-movement patterns for a pair of time series by the convolution technology. It also learns the temporal contextual relationships

Manuscript received December 1, 2012; revised August 26, 2015. Corresponding author: Lu Bai (email: bailucs@cufe.edu.cn).

between the evidence in different time windows through a BiLSTM process. By obtaining the co-movement attention with the captured co-movement patterns and temporal contextual relationships for stock time series, CAM focuses on the periods where two stock time series are strongly interdependent. CAM is supervised by the stock sector information since the stocks in the same sector may behave similarly during the trading process [9]. As the obtained co-movement patterns reveal the stock-level abnormality rules in different sectors, we propose a mechanism to fuse all the stock-level abnormalities to the market-level abnormalities in a bottom-up way. We verify the effectiveness of CAM on the real-world stock data and compare our method with state of the art methods on the market abnormality detection task.

In summary, the main contribution of this work including:

- We formalize and learn the interdependency rule between the stock time series as, the co-movement pattern, a high-dimensional co-varying relationship between two stock time series.
- We propose a neural network model CAM to automatically compute the co-movement attentions and the stock-level abnormality periods by utilizing the interdependency rules between pairs of stocks and the temporal contextual information of stock time series simultaneously.
- We propose a mechanism to detect the market-level abnormality periods by the obtained stock-level abnormality periods in a bottom-up way.
- Our model is supervised by the stock sector information, and we observe that CAM not only predicts the stock sector accurately but also discovers the co-movement attentions which can be used to detect both the stock-level and market-level abnormalities for the stock market that performs consistently with the real known financial crisis events.

II. Preliminary

In Economics, most related studies in stock market dynamics are related to the price, this work discusses the problem with the time series as the one-dimensional sequence. Note that this can be extended easily to the problems with the multi-feature time series. To explore the interdependency patterns between a pair of time series, we concatenate two aligned time series into a matrix which we define as the dyadic time series.

Definition 1: (**Dyadic Time Series**) A dyadic time series d is denoted as an $\mathbb{R}^{2\times T}$ matrix $d = [s_0, s_1]^T$, where each $s_i = (s_i(0), s_i(1), ...s_i(t), ...s_i(T))$ $(i = 0 \text{ or } 1)$ is an $\mathbb{R}^T$ vector which represents monadic time series (or a one-dimensional sequence) with T time windows.
In the practical applications, the monadic time series usually refers to the observation from a single source (e.g. the price time series from a given stock), and the dyadic time series records the observation about the interdependency for any pair of specific stocks.

We define the interdependency rule between a pair of monadic time series as a high-dimensional co-varying relationship as Definition 2.

Definition 2: (**Co-movement Pattern**) Given a length T dyadic time series d, the co-movement pattern is denoted as an $\mathbb{R}^{K\times 2\times T}$ tensor C, where K controls the dimension of the co-varying between the two related monadic time series of d.

After a convolution operation between the co-movement pattern and the raw dyadic time series, we obtain the evidence of the interdependency between two related time series. We leave the detail computation to the next section and define the co-movement attention as a vector which is computed based on the obtained evidence.

Definition 3: (**Co-movement Attention**) Given a length T dyadic time series d, the co-movement attention is denoted as an $\mathbb{R}^{T'}$ $(T' < T)$ vector $a = (a(0), a(1), a(2), ..., a(T'-1))$, where $a(i)$ $(0 \leq i \leq T'-1)$ refers to the co-movement attention value of d at the i-th window.

The co-movement attention records the evidence of interdependency relationship between two related time series. That is, the bigger value a_i has at timestamp i, the stronger interdependency between the two related time series (or the dyadic time series) in the i-th time window is. This value is summarized from the raw time series by CNN [10] and BiLSTM [11] module of our model. The computation process of co-movement attention also considers the temporal contextual information along the time dimension of the time series and thus the resulted attention values will not be overfitted to the local abnormalities. Since some operations in our model which will squeeze the original length of the dyadic time series into a dense representation that still preserves the original temporal order, the length of the co-movement attention $|a|$ or T' is an integer which is smaller than T. We leave the details to obtain T' in Section III.

As it is discussed in the introduction, during the market abnormality periods, the pairs of stock time series in a market are generally considered to be more interdependent than normal periods. Therefore, we define the concept, significant co-movement period, to describe and record this stock-level phenomenon.

Definition 4: **(Significant Co-movement Period)** Given the co-movement attention a, the significant co-movement period is a set of time windows I', for which $\forall i \in I'$, $a(i)$ is bigger than the attention values $a(j)$ $(\forall j \in I - I')$, where $I = \{0, 1, 2, ..., T'\}$ is the set of all the time window of a, $|I'|/|I| = \alpha$ and α $(0 < \alpha < 1)$ is the significant degree for the obtained significant co-movement period.

With the notations of co-movement attention and the significant co-movement period, our system can focus on different parts of dyadic time series with the significant degree α, e.g. we can set $\alpha = 0.01$ to make the system focus on the windows where the corresponding attention values are bigger than the values in the 99% remaining windows. Since the detection of the significant co-movement period is based on the co-movement attention, how to learn the co-movement attention is the first step for this work. We describe the problem to learn the co-movement attention as follows:

Definition 5: (**Co-movement Attention Learning**) Given a set of length T dyadic time series D with the corresponding sector label set L, $L(d)$ is the sector label of d ($\forall d \in D$), A is a set of the co-movement attentions for all the dyadic time series in D and a_d ($a_d \in A$) is the corresponding co-movement attention for d ($d \in D$). Then our goal is to estimate the optimal co-movement attention set A which generates the most approximate sector labels to the real corresponding sector labels. This problem can be formalized as:

$$\arg\min_{A} - \sum_{d \in D} p(L(d|A) \log(p(L(d))), \quad (1)$$

where $p(L(d))$ is the ground-truth probability to observe label $L(d)$ from the data and $p(L(d|A)$ is the probability to estimate d as with the label $L(d|A)$ by the model .

The reason to use the sector information as the supervision for this problem is that the sectors play an important role in affecting the co-movement patterns between the stocks [12] and the co-movement patterns tend to be similar for the stocks in the same sector or industry [13]. Since the co-movement attention is obtained based on the computation result of co-movement pattern and raw dyadic time series, the co-movement attentions of the dyadic time series from the same sector are also tend to be analogs.

What's more, this problem is an NP-hard combinatorial optimization as it requires enumerating different combinations of co-movement attention values for all the time windows, and we propose a deep learning model to solve it. We leave the details about the model and training process in Section III.

After solving the aforementioned problem in Definition 5, we obtain the co-movement attention for each pair of stocks in a market and thus can get the related significant co-movement periods. However, since the obtained significant co-movement period only reveals the stock-level abnormalities, we need to further analyze them in order to detect the market-level abnormalities. Therefore, we define the following notation to describe the market-level abnormality.

Definition 6: (**Market Abnormality Period**) Given a set of length T dyadic time series D for all the stock time series in a market, A is the corresponding co-movement attention set for D and the related significant co-movement period $I'(a)$ for $\forall a \in A$. Then the market abnormality period is a set of time windows I'', where for $\forall i \in I''$, the i-th window is in the significant co-movement periods, with a cutoff ratio β ($0 \leq \beta \leq 1$), of all the dyadic time series in D.

As we introduced in Section I, the co-movement between the stocks in a market will become widespread during the hazardous periods [1] and thus the market abnormality period can be defined as the period when a certain ratio (β) of dyadic time series are strongly co-movement. To get the market abnormality period, it is necessary to aggregate all the co-movement attentions in the dyadic time series set D for a market. We formalize the detection of the market abnormality period in the following.

Definition 7: (**Market Abnormality Period Detection**) Given a set of length T dyadic time series D for all the stock time series in a market, A is the corresponding co-movement attention set for D and the related significant co-movement period $I'(a)$ for $\forall a \in A$. The objective is to get the set of windows I'' which covers the real market abnormality periods with the stock-level abnormality periods $I'(a)$ for $\forall a \in A$.

The key to solving this problem is to get the aggregated abnormality periods which can represent the abnormality co-movement attentions of the dyadic time series with a certain ratio. We first apply a naive method to detect the market abnormality periods by counting the significant co-movement periods for the related time series. However, since this method faces the information loss problem by discarding too many insignificant results, we also propose a method to detect the market abnormality periods with Dempster-Shafer fusion from the evidence theory [14]. The Dempster-Shafer fusion mechanism allows our model to utilize the information from the data in the remaining $(1 - \alpha)$ windows when computing the significant co-movement periods. Since its fusion process is commutative, the aggregation using Dempster-Shafer fusion is also commutative: the aggregation process can be in any order without affecting the final result. We discuss the details in Section III and verify this method covers more real market abnormalities than the naive voting-based method in the experiment with more information.

The whole solution for Definition 5 and 7 results in our framework and the obtained market abnormality periods can also be used to evaluate the qualities of the learned co-movement attentions.

III. Our Framework

We propose the Co-movement Attention Model (CAM) to learn the co-movement attentions automatically. As it is shown in Figure 1, during the training process, the CAM model captures the co-movement evidences for pairs of stocks and the co-movement attention vectors by further utilizing the temporal contextual information along the time dimension. After the training of the CAM model, stock-level abnormality periods are obtained. Then, to explore the market-level abnormality periods, we propose the market abnormality detection module to detect the market abnormality periods by fusing all obtained stock-level abnormality periods.

A. Capture Co-movement Evidences for Pairs of Stocks

This process applies two convolution neural networks [10] and a max pooling [15] to capture the co-movement evidences for pairs of stocks. Its output is a $\mathbb{R}^T$ vector E which records the co-movement evidence between a specific pair of stocks. The kernel size or scan window for each related convolution or max pooling operation is N. The first convolution transforms the $\mathbb{R}^{2 \times T}$ input dyadic time series as the $\mathbb{R}^{K \times 2 \times T}$ tensor and its convolution kernel is the co-movement pattern for the specific pair of stocks. The second convolution transforms the $\mathbb{R}^{K \times 2 \times T}$ tensor obtained by the first convolution to a $\mathbb{R}^T$ vector as the output E. Therefore, this process tracks the K-dimension co-varying rules between the specific pair of stocks. All the convolution operations are the one-stride convolution

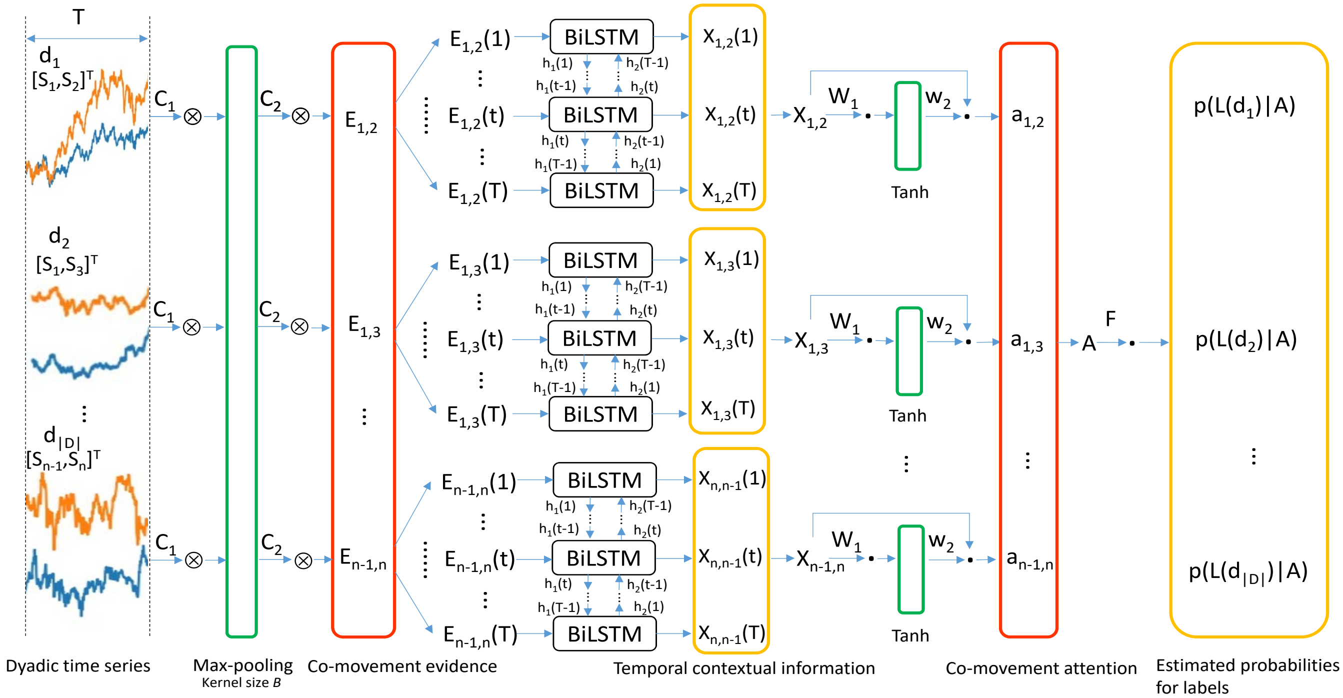


Figure 1. The framework of Co-movement Attention Model CAM. CAM captures both the interdependency patterns between pairs of stocks and the temporal contextual information along the time dimension. Supervised by the sector label information, it focuses on the stock-level abnormality periods automatically.

along the time dimension and this process is computed by Equation (2).

$$E = \max(d \otimes C_1 + B_1) \otimes C_2 + B_2, \tag{2}$$

where $\otimes$ is the convolution operator, $\max(*)$ is the max-pooling operation, B_1 and B_2 are the corresponding biases, C_1 and C_2 are the convolution kernels for the first and second convolutions shared by all the dyadic time series. Since both the convolution and max-pooling operations will change the length of the original time series, the length of the resulted vector E can be computed by the following equation.

$$|E| = T - 3(N + 1), \tag{3}$$

where T is the length of the original time series (e.g. $T = |S_1|$ or $|S_2|$), and thus E is an $\mathbb{R}^{T-3(N+1)}$ vector. Note that E collects the basic co-movement evidence between two time series in the same window and it has the same length as the final attention vector.

B. Capture the Temporal Contextual Information

With the basic evidence within the window for a specific pair of stocks, CAM further explores the temporal contextual information for pair of stocks between the time windows by a Bidirectional LSTM (BiLSTM) [11] with $2H$ hidden states. This BiLSTM represents E as a $\mathbb{R}^{|E| \times 2H}$ matrix X which incorporates the temporal contextual information along the time dimension and X is computed by Equation (4).

$$X = \begin{bmatrix} h_1(1), & h_1(2), & ...., & h_1(|E|) \\ h_2(1), & h_2(2), & ...., & h_2(|E|) \end{bmatrix}, \tag{4}$$

where $h_1(t)$ and $h_2(t)$ are the forward and backward hidden state values at time t respectively. They can be computed by the following equations.

$$\begin{aligned} i(t) &= \delta(W_{ii} \otimes E(t) + b_{ii} + W_{hi} \otimes h_x(t-1) + b_{hi}) \\ f(t) &= \delta(W_{if} \otimes E(t) + b_{if} + W_{hf} \otimes h_x(t-1) + b_{hf}) \\ g(t) &= tanh(W_{ig} \otimes E(t) + b_{ig} + W_{hg} \otimes h_x(t-1) + b_{hg}) \\ o(t) &= \delta(W_{io} \otimes E(t) + b_{io} + W_{ho} \otimes h_x(t-1) + b_{ho}) \\ c(t) &= f(t)c(t-1) + i(t)g(t) \\ h_x(t) &= o(t)tanh(c(t)) \end{aligned}$$

where $x = 1$ or 2, $E(t) \in E$, $\delta(*)$ is the sigmoid function, and $tanh$ is the hyperbolic tangent function. The W_{i*}s are the weights from the input to the different gates and the W_{h*}s are the weights between the hidden states and the different gates. All the b_*s are the corresponding biases. Since $h_1(t)$ and $h_2(t)$ are trained simultaneously, BiLSTM learns the contextual relationship between the different time windows in two directions. This bidirectional trait improves the model's ability to early warn the risk of abnormality before it actually happens. We provide detail discussion about this in the experiment section.

C. Generate the Co-movement Attention

With the $\mathbb{R}^{|E| \times 2H}$ matrix X of temporal contextual information, CAM generates the co-movement attention through a linear transformation method. As we introduced in Section II, the co-movement attention is a vector a which indicates the co-movement degrees based on the co-movement evidence. The length of a is the same as E ($|a| = |E|$), so that each element

$a(i)$ of a can be associated to a section of co-movement evidence E_i in a corresponding time window i ($i \in [0, |a|-1]$).

To get co-movement attention vector a, inspired by the self-attention in NLP [16], we design an $\mathbb{R}^{|2H|}$ vector m which records the contributions of the corresponding BiLSTM hidden states for the related co-movement attention and it can be obtained by the following equation.

$$m = w_2^T \cdot tanh(W_1 \cdot X), \tag{5}$$

where W_1 is an $\mathbb{R}^{f \times |E|}$ matrix, w_2 is an $\mathbb{R}^f$ vector and the parameter f is an integer constant which refers to the latent embedding dimension. With the hidden state contribution vector m, the co-movement attention vector a can be calculated in Equation (6).

$$a = softmax(X \cdot m^T) \tag{6}$$

This design allows our model to utilize all the temporal contextual information from the BiLSTM hidden states and get the correct co-movement attentions for the dyadic time series. We verify that this method can get the better co-movement attentions than the existing methods in the experiments.

D. Training with Sector Classification

As it is mentioned in Section II, the sector information affects the co-movement patterns between the stocks. Therefore CAM learns the co-movement attentions under the supervision of the sector label information. Suppose A is the set of all the obtained co-movement attentions, and we represent it as a $\mathbb{R}^{|D| \times |a|}$ matrix, where D is the set for all the dyadic time series and a is any obtained co-movement attention ($a \in A$). Then we use a fully connected layer to get the estimated probabilities of the sector label for the corresponding co-movement attention in A by Equation (7).

$$[p(L(d_1)|A), ..., p(L(d_{|D|})|A)]^T = A \cdot F, \tag{7}$$

where F is a $\mathbb{R}^{|a| \times |L|}$ matrix which transforms the co-movement attention values to the corresponding estimated probability distribution for all the labels in L. The parameters of our model are optimized by Equation (1) with the Adam optimizer [17] in the backpropagation way.

E. Market Abnormality Detection

Based on the proposed neural network model, we obtain the co-movement attention which represents the degree of co-movement for each pair of time series and thus can get the significant co-movement periods with the attentions. Since the obtained significant co-movement periods are only the stock-level abnormalities, how to use these stock-level abnormalities to generate the market-level abnormality periods is the next primary issue. This requires the modeling of the relationship between the stock-level abnormalities and market abnormalities for each window of the co-movement attentions. To this end, we propose two bottom-up methods, VOTE, and DSE, to detect the market-level abnormality based on the stock-level abnormality. While VOTE is based on the naive voting mechanism, and the DSE based on the evidence theory [14].

VOTE mechanism. VOTE follows on a naive idea that if more sectors vote for a time window as the abnormality period, then the market status in this time window will have higher possibility to be abnormal. It adds up the observation value of corresponding time windows for the top valued weights in each co-movement attention vector a ($\forall a \in A$). The details of VOTE is in Algorithm 1.

```
Data: The co-movement attention set A with significant degree
      α, cutoff ratio β.
Result: Market abnormality period I''.
begin
    I'' ← φ
    Initialize P as an R^|a| vector with all 0 elements
    for each a in A do
        /* stock-level abnormality periods   */
        I'(a) ←find the top α valued indices from a
        for ∀i in I'(a) do
            P[i] ← P[i] + 1              /* count the
            observations in the i-th window */
        end
    end
    /* market-level abnormality periods    */
    I'' ←get the top β valued indices from P
    Output I''
end
```

Algorithm 1: VOTE

Since the main computing process of VOTE relates to the numbers of the attention vectors $|A|$ and the number of the elements in $I'(a)$, the time complexity for VOTE is $O(|A| \times |a|)$. The problem of the VOTE mechanism is that it only uses the data in the significant co-movement periods above the significant degree α, and the most remaining observations are ignored by VOTE: it generates results with low coverage of the real market abnormalities.

```
Data: The co-movement attention set A with significant degree
      α, cutoff ratio β, the set I which contains all the
      window indices of A.
Result: Market abnormality period I''.
begin
    I'' ← φ
    Initialize P, P_0, P_1, and P_10 as R^|a| vectors with all 0
    elements
    for ∀a ∈ A do
        I'_t ←find the top α valued indices from a
        I'_b ←find the bottom α valued indices from a
        for ∀i ∈ I'_t do  P_1[i] ← P_1[i] + 1
        for ∀i ∈ I'_b do  P_0[i] ← P_0[i] + 1
        for ∀i ∈ I − I'_t − I'_b do  P_10[i] ← P_10[i] + 1
    end
    for i ∈ [0, |P|] do
        P[i] ←Dempster-Shafer fusion with P_0, P_1, and P_10
    end
    I'' ←get the top β valued indices from P
    Output I''
end
```

Algorithm 2: DSE

DSE mechanism. To make use of the information which is totally discarded by VOTE, we propose the DSE mechanism by applying the Dempster-Shafer fusion [14] from the evidence theory. As it is shown in Algorithm 2, DSE extends the significant degree α to determine both the top and bottom significant co-movement periods. Therefore, the market can be in one of the hypothesis abnormality statuses (“Yes”, “No”, or ”Unknown”) at each time window, where the “Yes” or “No” status of a window is decided by whether the attention value of that time window is in the top or bottom α ratio of the time windows for an attention vector, and the remaining time windows are in the “Unknown” statues. We provide the vector P_1, P_0, and P_{10} to record the frequencies of “Yes”, “No”, or “Unknown” in each window and get the prior probabilities with them for the corresponding statues at different time windows. Then, we use the Dempster-Shafer framework [14] to compute the posterior probabilities for the market abnormality periods at the corresponding time windows. We verify that the result of DSE better covers real market abnormalities than VOTE with more information in the experiment.

What’s more, to visualize the market abnormality period on the original time axis needs a process to convert the time windows of the market abnormality period into the originally covered range since the CNN changes the size of the window of the time series. According to the CNN structure in $III.A$, the size of the practically covered time window N' can be obtained by $N' = 3N - 2$, where N is the size of the original time window.

IV. Experiments and Discussion

A. Dataset

We compare our methods with other methods on the S&P 500[1] and NASDAQ 100 datasets [18]. The details are shown in Table I.

Property	S&P500	NASDAQ100
Company number	470	100
Time stamps	1,762	2,000
Original instants	851,264	200,000
Dyadic time series number	110,215	4,950
Practical instants	194,198,830	9,900,000
Sector combination number	66	21
Start date	2010-01-04	2016-07-26
End date	2016-12-30	2017-08-26

TABLE I
Dataset statistics

In order to learn the co-movement attentions between different pairs of stock time series and detect the market abnormality periods for the stock market, we enumerate and concatenate all the pairs of the stock time series in each dataset, and this results in the dataset containing over 100,000 dyadic time series (S&P 500). Since every dyadic time series has the same number of time stamps as the original time series, the final instant number for the resulted set of dyadic time series is over 100 million. The sector combination number refers to the different sector combinations for these dyadic time series, e.g. the sector combination could be “Energy and Health Care”, “Energy and Financial”, etc.

[1]https://www.kaggle.com/dgawlik/nyse

B. Experiment settings and benchmark.

Our experiment consists of four parts. First, we demonstrate the capability to learn the co-movement attention with CAM and other methods. Second, we compare the learned attentions from all the methods on the market abnormality periods detection task by using the market index data and the real-world financial abnormality events as the ground truths. Third, we analyze the training process of CAM with the sector classification task. Last but not least, we analyze the scalability of all mentioned methods.

Labeling methods. Since the sector classification task needs a sector label for each pair of dyadic time series, we label the dyadic time series according to the alphabetic order numbers of all the section combinations. e.g. 0 for a pair of time series which belongs to the sector “Energy and Financial”, 1 for a pair of time series which belongs to the sector “Energy and Health Care”, etc.

Implementation of CAM. We implement two versions of CAM-LCNN and CAM-BLCNN to show the different performances of LSTM and BiLSTM on our tasks. In the remaining part, we simplify CAM-LCNN and CAM-BLCNN to LCNN and BLCNN respectively for brevity.

Comparison methods. Our methods are compared with state-of-the-art methods which widely adopted in most related studies and a neural network model based on the original self-attention mechanism.

- Sliding Window Correlation (SWC) [19]. SWC is the classic method in many related studies, where the main idea of SWC is to divide the two related time series into sections with the same window size, and then calculate the Pearson correlation coefficient for each corresponding pair of the time series section in a window. The result of SWC is an attention weight vector for each window.
- Wavelet Coherence Correlation (WCC) [5]. Wavelet coherence is used in many recent works to study the inter-dependency patterns between time series. However, since most current wavelet coherence methods only give a heat map according to the wavelet coherence for each pair of time series which indicates the related correlation degree on each scales, it can not discover the co-movement attentions automatically. Therefore, we implement WCC based on the Morlet wavelet to compute the attention weight vector. WCC first calculates the wavelet coherence for a pair of time series. the wavelet coherence is an $\mathbb{R}^{J\times T}$ matrix which means the different correlation degrees on J scales. Then, WCC computes the sum along the J dimension, and get an $\mathbb{R}^{T}$ vector which refers to the correlation degree of each time stamp for the time series. In order to keep in accordance with other methods, we apply a sliding window framework to convert its result into an attention weight vector of the same size as other methods.

- SELF. We also implement a dyadic time series version of the original self-attention model [16] which could also be used to collect the co-movement attention vectors for comparison.

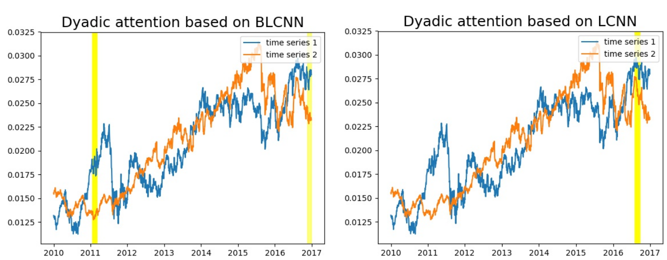
Ground truths for abnormality detection. Since Economic studies support that when the stock market declines the interdependency between related stock time series will become strong and widespread [20], we use the real stock market declines as the ground truth to measure the quality of the obtained co-movement attention which represents the co-movement degrees of the stocks.

Concretely, we provide two ground truths to verify the learned results by our methods. First, we collect the periods of the major stock market declines[2] from 2010.01 to 2016.12, and check whether or not the methods covered the related events. Second, we measure the market fluctuations and bubbles based on the market indices of "SPDR S&P 500 ETF Trust Index" (SPY) and INDEXNASDAQ (NDX) for all the methods.

We implement all the models in a prototype system which is based on the GPU version Pytorch[3], and thus they can be compared equally with the GPU environment. Our experiments are completed on a workstation with E3 CPU, 64 GB RAM, and Quadro P5000 GPU.

C. Co-movement Attention

To figure out the parts in which the related time series have strong interdependency patterns, we make our models focus on the top α ($\alpha = 0.01$ in this experiment) ratio of the attention values of the corresponding attention vector. We show the significant co-movement periods for a specific dyadic time series in the S&P 500 dataset in Figure 2, where we highlight the obtained significant co-movement periods as the yellow bars. From Figure 2, we can notice that our methods



(a) BLCNN (b) LCNN

Figure 2. Co-movement attention results (2010.1-2016.12)

show the capability to focus on some parts of the related two stock time series. Intuitively, this means the related stocks are interdependent strongly in the highlighted periods. To figure out whether or not these found results (or stock-level abnormalities) are related to the real market abnormalities, we provide experiments in the next section to compare the market-level abnormalities fused by the obtained co-movement attentions.

D. Market Abnormality Period Detection

In this experiment, We first compare the performances on detecting the market abnormalities with two fusion mechanisms (VOTE and DSE) on WCC and BLCNN, since WCC is the most efficient traditional algorithm and BLCNN is the best version of CAM. Then we compare the abilities of all mentioned methods to capture the market abnormality periods. The remaining experiment is twofold: to begin with, we compare the detected abnormality periods on the market index with several quantitative metrics for all mentioned methods, and then we illustrate the detection results on the real market crashes or bear market. For all the experiments in this part, the result for each deep learning model is obtained by the average after training and testing for 5 times independently and the cutoff ratio β is set to 0.03. Since the detected market abnormalities are based on the obtained co-movement attentions, the detection performance indicates the qualities of these discovered co-movement attentions.

VOTE v.s. DSE. We compare the coverage ratio of VOTE and DSE for WCC and BLCNN on the ground truth market declines (during 2010.01-2016.12). This comparison is conducted on the S&P 500 dataset. We list the result in Table II, and it shows noticeably that DSE can capture more ground truths than VOTE. This proves that the ignorant information of VOTE indeed helps DSE get better results. Therefore, all the rest comparisons are conducted with the DSE fusion.

Ticker number		25	50	100	200	470
WCC	DSE	**30%**	**40%**	**50%**	**40%**	**40%**
	VOTE	0%	0%	0%	20%	20%
BLCNN	DSE	**10%**	**40%**	**50%**	**80%**	**90%**
	VOTE	0%	0%	10%	20%	80%

TABLE II
VOTE V.S. DSE (COVERAGE RATIO) ON S&P 500 DATASET

Date	Decline	BLCNN	LCNN	SELF	WCC	SWC
2010.03-05	-16%	√	√	√	√	
2011.05-09	-19.4%	√	√	√	√	
2011.11	-9.8%	√		√		
2012.03-05	-9.9%	√	√			
2012.08-11	-7.7%	√	√	√		
2013.05	-5.8%		√			√
2014.01	-5.8%	√	√	√		√
2014.08	-7.9%	√	√			√
2015.05-08	-12.4%	√	√	√	√	√
2015.10-2016.01	-13.3%	√	√	√	√	√

TABLE III
COVERAGE OF THE MAJOR MARKET DECLINES ON S&P 500 DATASET

Coverage comparison of all methods. By specifying the DSE fusion mechanism, we compare the coverage ratio of each method on detecting the ground truth major declines for the stock market which consists of 470 stock time series of S&P 500 dataset during 2010.01-2016.12. This result is shown in

[2]https://www.cnbc.com/2018/01/16/us-stock-market-is-a-bubble-says-sovereign-wealth-fund-advisor.html
[3]https://github.com/hkharryking/comovement

Table III. We can observe that, with the same cutoff value, both BLCNN and LCNN cover the most (90%) real market declines than others; SELF cover 80% declines. We could also note that SWC only covers 50% declines, and this is because its local computational structure cannot capture the contextual information between the different windows. What's more, since WCC treats the information on each of its scales equally, it can not focus on the market-level abnormalities automatically. Therefore, it performs the worst in this task.

Quantitative comparison of all methods. We provide two metrics, the accumulative market fluctuation (AF) and bubble (AB), to further measure the abnormality periods quantitatively. To compute the two metrics, we merge all the continuous periods in the market abnormality period results to obtain a continuous period set W_c'', and then add up different metrics on the market indices within each period in W_c'' to get the accumulative result for each method. Generally speaking, the bigger the AF value is, the riskier the detected periods are; AB is the metric which reveals the risk of the "market bubble" [21] before the market crash or bear market. The detail for each metric is listed in the following.

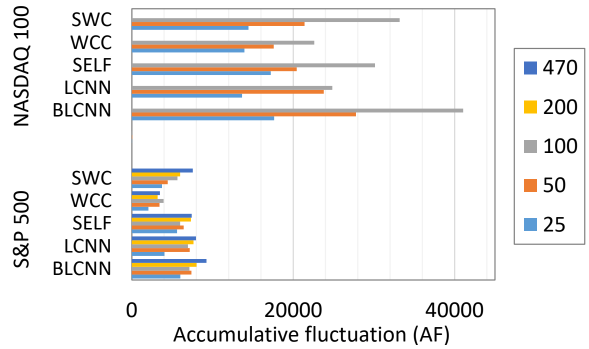
$$AF = \sum_{\forall i \in |W_c''|} \max(S_m, i) - \min(S_m, i), \quad (8)$$

$$AB = \sum_{\forall i \in |W_c''|} \max_{gain}(S_m, i), \quad (9)$$

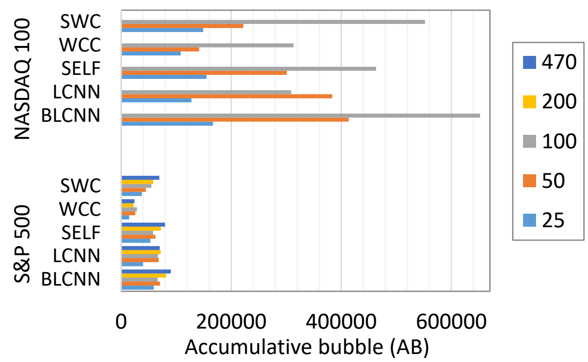
where function $\max(S_m, i)$ and $\min(S_m, i)$ return the biggest and smallest values of the series of market index S_m on the i-th period respectively, the function $\max_{gain}(S_m, i)$ return the maximum continuous gain of the series S_m on the i-th period. Note that, our method can be used to detect the abnormalities for any customized market or even the unknown market although we use the S_m as the ground truth for the market status in this work. We list the comparison results on all two metrics in Figure 3.

We can observe from the Figures 3, the BLCNN performs the best among all methods and we can observe that BLCNN can detect the market abnormality well even with a relatively small scale of data. This proves that the BiLSTM helps to capture the contextual information of the dyadic time series, and it is effective to utilize the latent information in Equation (5). We find that much to our surprise, the SWC beats the WCC on all the data scales. This is because the WCC method aggregates the latent information for all scale just in an equal way without further exploration. It is interesting that according to Figure 3 (b), BLCNN gets higher AB value than others and this shows its ability to discover the risky periods of the "market bubble" before the real market declines. This is very important to early detect the market abnormalities in the practical applications.

We also demonstrate the visual results of the abnormality detection of BLCNN on different data scales of S&P 500 dataset in Figure 4. It can be clearly seen that, with the increasing of the number of time series (from 25 to 470 stock time series), BLCNN can find the real-world abnormality periods more accurately. The result in Figure 4 (d), where the abnormality events are explicitly labeled, illustrates that BLCNN has good capability to capture the co-movement patterns, and it even captures the abnormality market status before some market crashes (e.g. 2015.05-08 crash) based on its learned co-movement attentions.



(a) AF results



(b) AB results

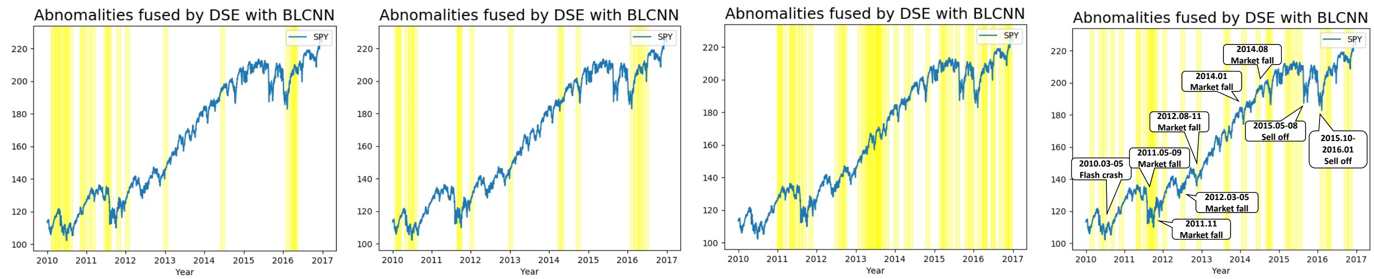
Figure 3. Quantitative comparison on AF and AB

E. Training of neural network models

This section demonstrates the effectiveness of stock sector classification using the proposed CAM model. Since traditional methods (WCC and SWC) compute the co-movement attentions without considering sector information, they are not included in this part. We select the data of 50 time series (that is 1,225 dyadic time series) in the period "1/4/2010 to 12/30/2016" for this experiment, where we sample 70% of the data for training and 30% of the data for testing. We use AUC to compare the classification performances over all sectors in different training epochs as shown in Table IV.

Epoch	BLCNN	LCNN	SELF
1	**0.5884±0.0032**	0.5597±0.0002	0.5041±0.0004
5	**0.6922±0.0010**	0.6262±0.0004	0.6081±0.0000
10	**0.7262±0.0008**	0.7234±0.0015	0.6552±0.0000
15	**0.7626±0.0013**	0.7399±0.0007	0.7085±0.0003
20	**0.7861±0.0000**	0.7533±0.0005	0.7311±0.0003

TABLE IV
COMPARISON ON CLASSIF. PERFORM. (AUC) ON S&P 500 DATASET



(a) 50 tickers (b) 100 tickers (c) 200 tickers (d) 470 tickers

Figure 4. Market abnormality detection by BLCNN on S&P 500 dataset

From the results in Table IV, we can find that the proposed CAM methods perform well on the classification task with the increasing of epoches. BLCNN performs the best among all the mentioned methods on the classification task. This result is also in accordance with the experiment in Section $IV.D$ where CAM captures the information about the market abnormality more accurately than other alternatives.

F. Scalability

As we implement all the methods on the GPU version Pytorch, we could discuss the scalability for all the methods mentioned in this work on the same hardware environment. Since the neural network models need training before detection, we first list and compare the training scalability for 15 epochs of the neural network models in Table V, and then we list the detection scalability for all the methods in Table VI. One can observe from Table V, LCNN is the most efficient neural network model in the training process since the size of hidden states in its LSTM layer is one half of the ones in BLCNN or SELF, and thus BLCNN and SELF take almost twice longer than LCNN during the training. As it is shown in Table VI, both the deep learning methods and WCC are efficient in the detection process, whereas SWC is the slowest method of all. This is because that SWC needs to compute the correlation in every window while other methods can extract the latent information from the whole time series.

Ticker num.	LCNN	SELF	BLCNN
25	20.68	42.22	42.99
50	93.12	191.28	193.20
100	382.56	781.14	806.89
200	1564.48	3222.99	3246.03
470	8682.95	18491.42	18239.78

TABLE V
COMPARISON ON TRAINING SCALABILITY (SEC.)

Ticker num.	LCNN	SELF	BLCNN	WCC	SWC
25	3.49	4.94	4.39	22.46	113.86
50	9.06	10.58	10.04	125.66	688.27
100	25.58	29.11	30.03	525.76	2905.44
200	89.64	104.48	107.17	2071.49	11791.04
470	474.33	565.36	555.36	11563.14	64594.44

TABLE VI
COMPARISON ON DETECTION SCALABILITY (SEC.)

V. RELATED WORKS

The interdependency between pairs of time series is a prevalent phenomenon in many fields such as stock market [22], recommendation system [23], Astronomy [24] or even molecular biology [25]. Current related methods primarily focus on analyzing the interdependency patterns between a pair of time series based on the pre-defined sliding window framework [1] such as the sliding window correlation methods [19] [4]. Since the correlation methods ignore the information beyond their pre-defined correlation framework, they are difficult to fully address the interdependency patterns between pairs of time series. Therefore, recent works also try to apply the wavelet transformation methods (many are the Wavelet Coherence, WC) [5] to explore more in-depth interdependency patterns for the related dyadic time series. However, although the wavelet transformation methods can compute the interdependency trends for time series on different scales respectively, they can hardly get an aggregated result which is related to the real market abnormality automatically. One solution is to provide the "heat map" [26] chart to show the interdependency patterns on different resulted scales for the time series and leave the identification works to human. Nevertheless, this is extremely labor-intensive for the human to monitor the abnormality status in a stock market with hundreds or even thousands of time series.

To further explore the relationship between pairs of time series, Lee et al. [27] concatenate the time series to a matrix and apply the convolution method to study the potential relationship with the resulted matrix. We follow this convolution method to track the interdependency patterns for pairs of stocks. Our work is the first to capture the stock-level abnormality by utilizing the interdependency patterns between pairs of stocks and the temporal contextual information along the time dimension simultaneously. Our work is also the first to detect the market abnormality based on the learned stock-level abnormalities in a bottom-up way.

VI. CONCLUSION

In this work we propose a neural network model, Co-movement Attention Model (CAM), to learn the co-movement patterns between pairs of time series and detect the market abnormality periods automatically. Our main contribution includes: we formalize the notations about the co-movement patterns and the co-movement attentions to address the interdependency rules between pairs of stock time series. We use the sector label information to supervise the learning process to get the co-movement attentions and patterns automatically. We detect the stock-level abnormality period with the obtained co-movement attention. Then, we propose a mechanism which is based on the Dempster-Shafer framework of the evidence theory to detect the market-level abnormality periods based on the stock-level abnormality periods without supervision. The experimental results on the real-world stock data verify that the sector label information is helpful to supervise the learning of the co-movement attentions, and the trained CAM model can also classify the dyadic time series to the correct sector label. The experiment about the market abnormality periods with the found co-movement attentions provides a quantitative way to compare the qualities of the co-movement attentions from all mentioned methods, and the result proves that CAM has captured the useful co-movement pattern which is related to the real market status. Besides, since our prototype system can solve the similar problem in any system which consists of numbers of dyadic time series, it could also be used to find out the abnormalities in any customized market.

VII. Acknowledgements

This work is supported by the National Natural Science Foundation of China (Grant No.61503422, 61602535), The Open Project Program of the National Laboratory of Pattern Recognition (NLPR), Program for Innovation Research in Central University of Finance and Economics, and Beijing Social Science Foundation (Grant No. 15JGC150). This work is also supported in part by NSF through grants IIS-1526499, IIS-1763325, and CNS-1626432, and NSFC 61672313.

References

[1] K. J. Forbes and R. Rigobon, "No contagion, only interdependence: Measuring stock market comovements," *The Journal of Finance*, vol. 57, no. 5, pp. 2223–2261, 2002.

[2] T. Conlon, H. Ruskin, and M. Crane, "Cross-correlation dynamics in financial time series," *Physica A: Statistical Mechanics and its Applications*, vol. 388, no. 5, pp. 705–714, 2009.

[3] Y. Le Pen and B. Svi, "Futures trading and the excess co-movement of commodity prices*," *Review of Finance*, vol. 22, no. 1, pp. 381–418, 2018.

[4] P. Banerjee, P. Yawalkar, and S. Ranu, "Mantra: A scalable approach to mining temporally anomalous sub-trajectories," in *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '16. New York, NY, USA: ACM, 2016, pp. 1415–1424.

[5] J. C. Reboredo, M. A. Rivera-Castro, and A. Ugolini, "Wavelet-based test of co-movement and causality between oil and renewable energy stock prices," *Energy Economics*, vol. 61, pp. 241 – 252, 2017.

[6] G. E. A. P. A. Batista, E. J. Keogh, O. M. Tataw, and V. M. A. de Souza, "Cid: an efficient complexity-invariant distance for time series," *Data Mining and Knowledge Discovery*, vol. 28, no. 3, pp. 634–669, May 2014.

[7] N. Laptev, S. Amizadeh, and I. Flint, "Generic and scalable framework for automated time-series anomaly detection," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '15. New York, NY, USA: ACM, 2015, pp. 1939–1947.

[8] L. C. D. L. Guansong Pang, Longbing Cao and H. Liu, "Sparse modeling-based sequential ensemble learning for effective outlier detection in high-dimensional numeric data," in *The 32nd AAAI Conference on Artificial Intelligence*, ser. AAAI '18, 2018.

[9] L. T. Junjie Chen and B. Zheng, "Agent-based model with multi-level herding for complex financial systems," *Scientific Reports*, vol. 5, no. 8399, 2015.

[10] M. Hu, Z. Li, Y. Shen, A. Liu, G. Liu, K. Zheng, and L. Zhao, "Cnn-iets: A cnn-based probabilistic approach for information extraction by text segmentation," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, ser. CIKM '17. New York, NY, USA: ACM, 2017, pp. 1159–1168.

[11] P. Wang, Y. Qian, F. K. Soong, L. He, and H. Zhao, "Learning distributed word representations for bidirectional LSTM recurrent neural network," in *NAACL HLT 2016, The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego California, USA, June 12-17, 2016*, 2016, pp. 527–533.

[12] A. Rua and L. C. Nunes, "International comovement of stock market returns: A wavelet analysis," *Journal of Empirical Finance*, vol. 16, no. 4, pp. 632 – 639, 2009.

[13] L. K. C. Chan, J. Lakonishok, and B. Swaminathan, "Industry classifications and return comovement," *Financial Analysts Journal*, vol. 63, no. 6, pp. 56–70, Nov 2007.

[14] C. Zhou, X. Lu, and M. Huang, "Dempster-shafer theory-based robust least squares support vector machine for stochastic modelling," *Neurocomputing*, vol. 182, pp. 145–153, 2016.

[15] C. Szegedy, W. Liu, and Y. Jia, "Going deeper with convolutions," in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*, 2015, pp. 1–9.

[16] Z. Lin, M. Feng, C. N. dos Santos, M. Yu, B. Xiang, B. Zhou, and Y. Bengio, "A structured self-attentive sentence embedding," in *International Conference on Learning Representations 2017 (Conference Track)*, 2017.

[17] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014.

[18] Y. Qin, D. Song, and H. Chen, "A dual-stage attention-based recurrent neural network for time series prediction," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, 2017, pp. 2627–2633.

[19] Y. He and A. Kusiak, "Performance assessment of wind turbines: Data-derived quantitative metrics," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 1, pp. 65–73, Jan 2018.

[20] L. Yarovaya and M. C. K. Lau, "Stock market comovements around the global financial crisis: Evidence from the uk, brics and mist markets," *Research in International Business and Finance*, vol. 37, pp. 605 – 619, 2016.

[21] A. Gerow and M. T. Keane, "Mining the web for the "voice of the herd" to track stock market bubbles," in *IJCAI 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16-22, 2011*, 2011, pp. 2244–2249.

[22] Q. Li, Y. Chen, J. Wang, Y. Chen, and H. Chen, "Web media and stock markets : A survey and future directions from a big data perspective," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 2, pp. 381–399, Feb 2018.

[23] Y. Sun, N. J. Yuan, X. Xie, K. McDonald, and R. Zhang, "Collaborative nowcasting for contextual recommendation," in *Proceedings of the 25th International Conference on World Wide Web, WWW 2016, Montreal, Canada, April 11 - 15, 2016*, 2016, pp. 1407–1418.

[24] M. Finn, E. E. Mamajek, K. Luhman, and S. Murphy, "New Low-Mass Wide Companions to Members of the Sco-Cen OB Association," in *American Astronomical Society Meeting Abstracts #229*, ser. American Astronomical Society Meeting Abstracts, vol. 229, Jan. 2017, p. 344.14.

[25] R. H. S. Brian P. English, "A three-camera imaging microscope for high-speed single-molecule tracking and super-resolution imaging in living cells," pp. 9550–9556, 2015.

[26] A. O. el Alaoui, G. Dewandaru, S. A. Rosly, and M. Masih, "Linkages and co-movement between international stock market returns: Case of dow jones islamic dubai financial market index," *Journal of International Financial Markets, Institutions and Money*, vol. 36, pp. 53 – 70, 2015.

[27] J. B. Lee, X. Kong, Y. Bao, and C. M. Moore, "Identifying deep contrasting networks from time series data: Application to brain network analysis," in *Proceedings of the 2017 SIAM, Houston, Texas, USA, April 27-29, 2017.*, 2017, pp. 543–551.