Granger causal weather time series forecasting simulation combined with mutual information

Fengkui Xu, Shibao Sun*, Pengcheng Zhao, Shaoyong Jia
College of Information Engineering, Henan University of Science and Technology, Luoyang, 471023, China
sunshibao@haust.edu.cn

Abstract. Meteorological science is becoming more and more familiar. The internal driving relationship of weather, timing sequence and its accurate prediction are hot topics in the field of meteorology. This paper proposes a granger weather, timing sequence causal prediction method combining mutual information. In this paper, the information entropy characteristic matrix of the Granger feature set is calculated through mutual information. It verifies the causal relationship of the weather, timing variables such as temperature, humidity and wind speed, it mines the effective input variables to achieve the purpose of reducing the dimension of time-series data features and improves the prediction accuracy of the target model. Finally, the open source weather data set is used to make a comparative experiment between qualitative and quantitative analysis of the proposed algorithm. The simulation results show that the method is feasible and effective.

1. Introduction
With the rapid development of artificial intelligence technology[1], more and more related technologies are applied to weather change, and many scholars have made many achievements in the driving relationship and predictability of weather change. Zhang Yugui[2] et al. put forward a model combined with radial basis function by virtue of the advantages of neural network in effectively processing nonlinear data. The main application scope of this kind of literature is nonlinear multivariate time series, which has high computational complexity and does not consider the internal response relationship of weather. Ma Qingtao[3] et al. found that the average relative error of air quality index simulation prediction results in the next seven days is 0.01 by establishing the prediction model of least squares support vector machine optimized by particle swarm optimization. Duen [4] et al. used LSTM neural network model and XGBoost ensemble learning method to predict PM2.5 for two periods. Although this kind of method considers weather data such as wind direction and wind speed to drive the internal response relationship, it is easily affected by other uncertain weather information, which leads to problems such as over-fitting of the resulting curve and affects the accuracy of the result.

In view of this, this paper will combine causal interpretation to analyze the dynamic mechanism of weather time series. It proposes Granger causal weather time series prediction algorithm combined with mutual information. Mutual information is introduced into the original Granger causal feature set, which makes the time series feature set solve the features with high correlation. The significance of causality statistics is tested by F test, and the causal matrix of weather time series data is described.
2. Granger causality and mutual information

The basic idea is that for two time series data[5], if one of the target time series Y introduces the historical information about the other time series X. The prediction accuracy of Y is higher than that of the target sequence Y only by the historical information of Y, then X is the Granger cause of Y, that is, X can help explain the future trend of Y. The regression formula for building Y is shown in (1), and the new regression formula after introducing X is shown in (2). If the predicted effect of formula (2) is better than that of formula (1), X is the Granger dependent variable of Y.

\[ Y_p(t) = \sum_{i=1}^{L} a_i Y_{t-i} + \epsilon_1 \]  

\[ Y_p(t) = \sum_{i=1}^{L} a_i Y_{t-i} + \sum_{k=1}^{K} b_k X_{t-k} + \epsilon_2 \]  

\[ Y_p(t) \] is the predicted value. \( a_i \) and \( b_k \) are the regression coefficient of the model. \( L \) and \( K \) are the order of the model. \( \epsilon_1 \) and \( \epsilon_2 \) are the prediction residual of the model.

Compared with Granger causality, the advantages of mutual information method are as follows:(1) It can express the nonlinear and linear relationship between variables;(2) It can indicate the correlation information between variables and has no other requirements on the distribution type of data variables. It describes the degree of information content according to the numerical form. It can indicate the flow of information related to the variables.

3. Granger causal time series prediction algorithm based on mutual information

3.1. Algorithm analysis and description

The Granger cause time series prediction framework combined with mutual information constructed in this paper is shown in Fig1.

![Fig.1 Granger causal time series prediction framework diagram based on mutual information](image)

In order to analyze weather change and its predictability, this paper proposes Granger causality, time series algorithm based on mutual information. It analyzes the indicators of causality between weather time series, and establishes a prediction model for this causality. The basic steps of the algorithm are as follows.

1. The first part of it is to perform data preprocessing operations. It needs to standardize the weather time series data and judge whether the data is smooth. And it performs augmented Dickie-Fuller (ADF) stabilization transformation on non-stationary time series. The time series, Granger test needs to be smoothed.
2. It begins the core process of the algorithm. The weather multivariate time series (MTS) has a higher dimension after vectorization. Firstly, Granger causality coefficient between MTS series is calculated by Granger causality regression method. The causality matrix is constructed and the Granger causality feature set is formed. Then, the mutual information method is used to further select the features of the original Granger causality feature set. And the original causal features are screened and filtered through the measurement of mutual information correlation dependence between sets, so as to remove the weak correlation feature set, retain the strong correlation feature value. Then it describes the different information flows of the causal relationship for Granger sequence prediction. The steps of the mutual information feature selection method are as follows. It first calculates the mutual information between the original Granger causality feature set, and obtains an $n \times n$ mutual information matrix $Z_{2}(X, Y)$. It then sets the correlation threshold $\theta_{1}(0 \leq \theta_{1} \leq 1)$, and selects the input dimensions that satisfy the correlation constraint formula $Z_{2}(X, Y) \geq \theta_{1}Z_{2}(Y, Y)$ according to it. It finally lets $Z_{2}(X, Y)$ sorted in descending order to get set $E$. It initializes the optimized subset of variables: $S=E$. When the variables $\{S_{i}\}, i \in (1,2, \cdots, n)$, it can remove weakly correlated feature sets such as the mutual information between input variable $X$ and output variable $Y$ ($Z_{i} = Z_{i}(S - \{S_{i}\}, Y)$). It then sets the redundancy threshold $\theta_{2}(0 \leq \theta_{2} \leq 1)$ to remove redundant variables, and its constraint formula is $Z_{2}(E, Y) - maxZ_{1} < \theta_{2}Z_{2}(Y, Y), S = S - \{S_{i}\}, \theta_{i} = argmaxZ_{1}$. It repeats the above steps to remove weakly related features and redundant features. When the number of selected strongly correlated variables reaches the hypothesis, the method ends.

Therefore, it proposes a predictive model that combines Granger causality. It makes both sides of the formula (1) (2) multiplied by $n$, as shown in formula (3).

$$\begin{align*}
Y_{p}(t) &= n \sum_{i=1}^{d} a_{i}X_{t} + n\varepsilon_{1} \\
Y_{p}(t) &= n \sum_{l=1}^{d} a_{l}Y_{t-l} + n\varepsilon_{2}
\end{align*}$$

(3)

It is added to a common item separately and proposes this common item, as shown in formula (4).

$$Y_{p}(t) = n \sum_{l=1}^{d} a_{l}X_{t} + \sum_{l=1}^{d} a_{l}Y_{t-l} + n\varepsilon_{1} = \sum_{l=1}^{d} a_{l}X_{t-l} + n\varepsilon_{2}$$

(4)

And it calculates $\sum_{i \in \{1,2, \cdots, n\}} Y_{p}(t) = n \sum_{i \in \{1,2, \cdots, n\}} Y_{p}(t)$. The $Y_{p}$ is a new Granger prediction model of $Y_{p}$. The formula is shown in (5).

$$Y_{p} = \sum_{i \in \{1,2, \cdots, n\}} \left(a_{i}X_{t} + \sum_{k=1}^{K} b_{k}X_{t-k}\right) + \varepsilon$$

(5)

3. If we assume a sequence $X$ It's another sequence of terms $Y$. F-test can be used to test whether the causal relationship has statistical significance. The test is based on the regression set model of the formula. It allows features with a certain causal strength as the input dimension of Granger prediction model($Y_{p}$. And it describes different information flows between MTS.

4. Granger prediction of weather multivariate data is carried out for matrix sequence with causality. It dependent variables and lag observation period (lag) required by input vector regression prediction model, and the prediction results are output.

4. Test Results and Discussions

4.1. overall composition of the experiment

Weather is an open source data set of UC1 on a certain city weather. The weather data set has a long time span, with 5717 sample size and 23 characteristic numbers. It is a complex multiple time series data set, which meets the experimental requirements. The simulation is run in win10 system (CPU Intel corei5-4210u / 4gram) with the Python language.

In the first group of experiments, the weather data set is used to prove the feasibility of the method, and the information flow direction is described and the corresponding prediction results are obtained. The second group of experiments shows the effectiveness of the algorithm, setting 80% of the data set as training data. 20% is the test set. And it compared with reference [2] about radial basis function, literature [3] about support vector machine and Literature [4] about long and short term memory
algorithm. Then it summarizes the experimental results. The third group of experiments as follows. It is to further verify the effectiveness of the algorithm in this paper, making it compared with other algorithms. The historical data length is 240 and the predicted length is 24.

4.2. evaluation method of experimental model
In order to better test the superiority of the algorithm and the experimental results, the root mean square error (RMSE) and mean absolute error (MAE) are used to evaluate the predictive effect. They can be used to judge the prediction accuracy from different angles. RMSE represents the square of the deviation between the predicted value and the true value and the number of observations. The square root of the ratio is used to judge the deviation between the predicted value and the true value. MAE is the average absolute error value. It shows the actual trend of prediction error.

4.3. Analysis of experimental results
The first group of experiments as follows. The ADF test results are less than the critical value of 1% significance level and the P value is less than 0.05. It indicates that the weather series is stable and can be tested by Granger causality test.

The weather original Granger feature set is calculated. The strong correlation feature set is retained after further screening and filtering by the mutual information method. The Granger causality is tested by F test. If the verification probability P value is less than the original confidence level of 0.05, there is a causal relationship between the two weather variables. Fig. 2 is the weather time series causality matrix after the F-test. The horizontal and vertical coordinates here correspond to the characteristics of nine strongly related variables of mutual information screening and filtering. The values of the causal matrix correspond to Granger causality index between weather time series variables. And the values of the diagonal in Fig. 2 express their own causality measures.

![Fig.2 Weather time series causality matrix after F test](image)

According to the weather time series causality matrix in Fig 2. There is a lot of Granger causality information flowing in, in which weather Event is target sequence data. According to the target variable Event with a large amount of Granger causal information inflow expressed in Fig.2. The number of effects predicted dependent variable features is 7, which is less than the number of variables in the original data set. Then it selects the maximum lag order lags=15 and input it into the algorithm prediction
The prediction result is shown in Fig. 3. The MI-causality has a good fitting degree. It proves the feasibility of the method.

![Fig. 3 The MI-causality prediction results](image)

The second group of experiments as follows. When different causality algorithms are used to build predictive models, lags=15 is used to verify and compare the efficiency of the algorithms. Different causal relationships are established by the algorithms proposed in literature[2], literature[3] and literature[4], and then the causality prediction model is established for the target variable Event. In order to better evaluate the actual performance of the weather prediction model, 80% is set as training data and 20% is set as a test set for the weather quantitative prediction experiment. And three indexes (Humidity, Mean Wind and Sea Level Pa) are evaluated respectively. The results are shown in Tab.1~Tab.3.

| Tab.1 Comparison of (Humidity) prediction results | Tab.2 Comparison of (Mean Wind) prediction results |
|-----------------------------------------------|-----------------------------------------------|
| **Type** | **RMSE** | **MAE** | **Times** | **Type** | **RMSE** | **MAE** | **Times** |
| RBF-causality | 36.82 | 31.65 | 67.56 | RBF-causality | 5.17 | 4.51 | 35.99 |
| SVM-causality | 33.40 | 27.81 | 57.34 | SVM-causality | 4.09 | 3.36 | 33.67 |
| LSTM-causality | 27.10 | 21.47 | 53.92 | LSTM-causality | 4.05 | 3.32 | 28.12 |
| MI-causality | 26.59 | 21.47 | 50.21 | MI-causality | 3.28 | 2.60 | 25.60 |

From the quantitative analysis results in Tab.1~Tab.3, it can be seen that the causal prediction MI-causality in this paper is better than other algorithms in the above three evaluation criteria. And one is equal to the LSTM-causality and one is slightly worse than the LSTM-causality.

The third group of experiments as follows. The historical data length is 240 and the predicted length is 24 on the weather data set. The MI-causality and LSTM algorithm respectively compare and evaluate the three indexes (Humidity, Mean Wind, Sea Level Pa). And the simulation results are shown in Fig. 4~Fig. 6. As shown in it, four indexes are better than LSTM algorithm, one is equal to LSTM algorithm, and one is less effective than LSTM algorithm. For long-term data, this algorithm is more stable, and most evaluation indexes are more accurate. That is to say, Granger time series causal prediction algorithm has low computational complexity under the constraint of strong correlation combined with mutual information. It proves the correctness of the theory and formula derivation.
5. Conclusion
In this paper, a Granger causality weather time series forecasting algorithm combined with mutual information is proposed. The algorithm uses mutual information to deal with nonlinear relationship and can help Granger causality indicate its information flow direction. And mine the dependent variable of effective prediction to reduce dimension. It show that the algorithm is feasible and effective for the prediction model established by complex multivariate weather data. My future work will refine the flow of detailed causal information.

Acknowledgments
This work was financially supported by National Natural Science Foundation of China (51474095), Henan Key Research Project (152102210277), Henan University Science and Technology Innovation Team Support Program Project (17IRTSTHN010), Henan University of Science and Technology Science and Technology Innovation Team Project (2015XTD011) and Henan University of Science and Technology Major Industry-University-Research Cooperation Cultivation Fund Project (2015ZDCXY03). * is the corresponding author of this article.

References
[1] Ma Leiming. Advances in artificial intelligence technology in weather forecasting [J]. Earth Science Progress,2020,35(06):551-560(in Chinese).
[2] Zhang Yugui,Wang Yi, et al. Weather forecast model based on radial basis function neural network [J].Journal of Guizhou University (Natural Science Edition),2018,35(01):69-72+103(in Chinese).
[3] Ma Qingtao, Shang Guobei. Haze prediction model of least square support vector machine based on particle swarm optimization [J]. Journal of Hebei University of Geosciences, 2019, 42(02): 51-55 (in Chinese).

[4] Duen-Ren Liu, Shin-Jye Lee, Yang Huang, et al. Air pollution forecasting based on attention-based LSTM neural network and ensemble learning. 2020, 37(3): n/a-n/a.

[5] Ren Weijie, Han Min. Summary of Research on Causality Analysis of Multivariate Time Series [J]. Acta Automatica Sinica, 2019: 1-15 (in Chinese).