Interactive Visual Analytics System for Paleoclimate Causality

Jung-Chen Chang National Taiwan Normal University

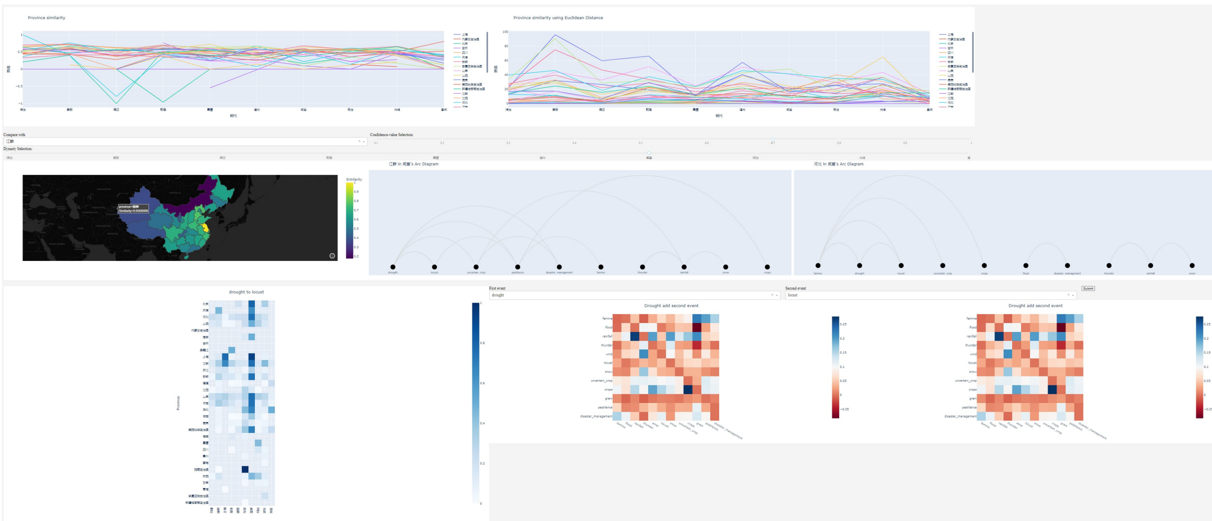

Figure 1: Interface of our interactive visual tool

Abstract

Climate data can provide valuable information. Understanding and identifying the correlations between climates is a crucial task in climate analysis, particularly when studying ancient climates, which is a topic of great interest to many meteorologists. However, data on ancient climates can only be collected from historical records or literature, which makes studying ancient climates more difficult. In this study, we use the REACHES (Reconstructed East Asian Climate Historical Encoded Series) dataset. We designed an approach based on the Apriori algorithm to discover frequently occurring associations among climate events. Based on these associations, we developed an interactive visualization tool that allows experts to explore the historical climate of the Ming and Qing dynasties spanning 600 years. This tool helps them understand the interactions between climates, enabling them to investigate the relationships between events and undertake more advanced research tasks, such as reconstructing historical climates or comparing the impacts among different climates.

1 Introduction

Climate science is the field of research that studies the Earth's climate system and its changes. It studies how the climate forms and changes, and its effects span many fields from agriculture to the development of human civilization [9, 10, 20]. Climatologists typically focus on studying climate patterns, events, and trends to better understand how the climate system works. Climate, however, is a long-term regional weather pattern influenced by various interacting factors. To understand its current trends and make predictions about the future, it is necessary to figure out the differences between different periods. A simple example is that climatologists may not be able to figure out anomalies just by looking at temperatures over the past decade. But when compared with long-term historical temperature data it becomes noticeable that the Earth is warming. It can be seen that comparing the climate of different periods is an important practice for climate study. However, there is no complete and systematic measurement of climate data for the past few hundred years, which leads climatologists to find other ways to collect climate information in the old time, such as words mentioned in the literature, ice cores from glaciers, or tree rings [6, 16, 21].

Although climatologists cannot have the climate measurements, such as temperature or precipitation, from several hundred years ago, an alternative way to infer the climate information from hundreds of years ago is to study the historical events, such as famine, drought, locusts, flood, etc. Compared to climate measurements, we have abundant event records in books, such as dynasties' history literature or emperors' biography. Some climatologists put effort into collecting massive amounts of books, transforming the text contents into organized datasets, and reconstructing ancient climate-related information to facilitate climate research [7, 13, 19]. REACHES (Reconstructed East Asian Climate Historical Encoded Series) is a dataset compiled from ancient literature that records various climate events and disaster events. The dataset includes abundant events that happened in the region of China from 1368 to 1911. Climatological domain experts have conducted multiple studies on top of the REACHES dataset [8, 15, 24]. For example, studying drought could cause what events or which local regions or time spans have significant differences in drought or rainfall severity.

Experts are often interested in discovering and explaining the causality under given conditions or looking for abnormal event relations when compared with other spatial regions and time spans. One common practice to discover causality is first exploring relations between events, then finding more evidence to verify the hypothesis. However, if the search space is too large, experts can only try a few combinations and miss the opportunity to discover "surprise" scientific outcomes. In REACHES dataset, even if experts only define the spatial granularity at the province level (more than 30 provinces), time granularity by emperors (26 emperors), and the coarsest event categories (13 events), there are still more than 10000 relations of combinations to explore. In addition, experts also require to study "unsymmetric" relations between events to avoid missing valuable combinations for further study. A naive but classic example of the unsymmetric relation between events is thunder and rain. We know that there must be causality between rain and thunder. Although thunder almost comes with rain, it might not have thunder when it rains. Because the classic correlation equation does not consider the direction between events, it could give a low value for the relation between rain and thunder. If so, experts could miss the opportunity to explore the causality. Therefore, comprehensively exploring the dataset and finding valuable combinations for future study is not trivial.

To assist experts in identifying potential hidden relations from large amounts of data, we deploy association rule learning [] to analyze relations of event combinations in the dataset. We also closely work with domain experts to develop an interactive visual system based on association rule learning results. The association rule learning is often used in large transaction data analysis. For example, frying pan, oil $\rightarrow$ egg is a rule found in a supermarket dataset, which means a customer often also buys eggs when the customer purchases a frying pan and oil. The found rules can be connected to discover more complex relations. We transform our records in REACHES dataset into transaction-like format. The events that happened in a small local region and a short time window are put together as a record. Then, all records are analyzed by associate rule learning to determine the strength of events' relations. The visual analytics tool cooperates multiple views and interactions to assist experts in exploring spatial regions and time spans with similar or dissimilar association rule set, comparing the detailed difference between two association rule set, and examining whether any additional event could strengthen the relation between a known events' relation. Users can interactively explore the huge amount of association rules, formulate hypotheses and verify them. We demonstrate our system through two use cases. The first use case is experts find that there exists some combination of events in a specific region and short time spans that are unexpected when using the correlation equation. The second use case is comparing the historical background of different periods through the spatial variation of association rules We also get user feedback from experts to discuss the advantages and limitations of our visual system.

The contributions of this study are as follows:

(1) Design interactive visualization tools to facilitate experts to observe ancient climate data.

(2) Find out the association rule that is unexpectable in specific time periods and spatial contexts.

(3) Through the association rule change on the space to

2 Related work

2.1 Climate analysis

Climate analysis is a complex field of study in which climatologists strive to clarify the interactions between different climate events and their impacts. To achieve this goal, the research of climatologists can be divided into several directions. The first is the simulation and prediction of the global climate model, through the long-term change trend to improve the accuracy and reliability of the climate model, allowing for better simulation and prediction of future climate changes [12,22,30]. Song et al. used neural networks to model future temperature changes and predict future summer temperature changes [23]. Mai et al. introduce a new framework to predict the spread of dengue outbreaks using data from extreme climate events [17]. The second direction focuses on analyzing specific climates to observe their patterns and impacts, aiding climatologists in understanding the mechanisms behind their formation and exploring the effects of climate change on ecological structures and societal factors [5]. Lin et al. used time-series analysis to identify periods of severe drought during the Qing Dynasty and focused on finding links between drought and other ecological and social variables [14]. Tian et al. Reconstructed the 1910 long-term series of migratory locust outbreaks in China based on historical documents, and indicated that there was a significant correlation between locust numbers and precipitation and temperature indicators [26]. In this study, compared to the previously mentioned studies that focus on the in-depth analysis of the correlation of a single combination, our goal is to calculate the correlation between different climate events, by allowing users to compare the Correlations, helping experts discover hidden climate combination relationships, and then use their expertise to do more follow-up research.

2.2 Spatiotemporal analysis

With the advancement of technology, we are now able to analyze larger and more complex datasets, and spatiotemporal analysis is an important research direction. By examining the distribution of events in different times and spaces through spatiotemporal data and geographic information, we can uncover hidden correlations between events and predict future spatiotemporal patterns [4, 18]. For example, the same type of climate can occur in different places, but the extent of its impact can vary greatly due to different geographical environments or spatiotemporal contexts. The key goal of spatiotemporal analysis is to discover implicit patterns and trends in such spatiotemporal data and extract valuable information from them. This type of analysis is applied in various fields. Anwar et al. modeled the congestion of urban road networks to effectively capture their spatiotemporal evolution [2]. Wang et al. provided a visual analytics framework to help users process large geographic data and their topological network relationships [29]. In this paper, we design a visualization tool to examine whether there are changes in the correlations between climates in different provinces or dynasties.

2.3 Correlation

Correlation is a very important concept in statistics [3, 25], which is used to describe the relationship between two or more variables. Correlation analysis can help us understand and explain the relationship between variables and gain a deeper understanding of the patterns and trends behind the data. Therefore, no matter what field of research, experts will try to find the correlation between different events. In modern data analysis, correlation coefficient analysis is a common method. In meteorology, for example, Vassoler et al. calculated the correlation between temperature and humidity and showed that the data are affected by a seasonal component [27]. Wang et al. studied the statistical characteristics of cross-correlation in the US stock market through the method of Pearson correlation coefficient [28]. In addition to these statistical methods, there are now many algorithms that discover co-occurrence patterns and dependencies between items by exploring frequent item sets and association rules in datasets [1, 11]. A classic example is that if a customer buys diapers, they are likely to buy beer as well. In this study, because most of the events in the data set can only know the number and frequency of their occurrence, we choose frequent pattern mining to find the correlation between events.

3 Task and Challenge

After interactively discussing with domain experts, we understand the experts' data analysis requirements and challenges, define the design goals the tool, and determine the tasks that should be completed to serve the goals. We are going to introduce them in this section.

3.1 Challenges and Goals

C1: Relation evaluation metrics experts used often give a low response on some event combinations that are known that they should have strong relation, such as the thunder and rain example introduced in Section 1 because their relations are not symmetric. This could result in missing interesting phenomena.

Attribute Name	Description
year_lunar_st	Start year of the subrecord/event(s) in lunar calendar
place_provin	Provinces
place_longit	Assigned longitude represented by the location of the city hall
place_latitu	Assigned latitude represented by the location of the city hall
event_code	Events are divided into four domains. Each domain has a categorizing structure from main category, subcategory, to vocabulary.

Table 1: Parameter names and explanations used in this study

C2: Experts know that the high strength of relations of some events could propagate from one location to another location over time. However, observing and discovering this type of trend is not trivial in a large spatial-temporal dataset.

C3: Some interesting or abnormal event relations only appear in specific locations and time windows. However, finding out these relations from a large number of event combinations, time intervals, and spatial location is not an easy task.

G1: Deploy a data analysis algorithm that can calculate the strength of relations of events on bi-directions.

Calculate and analyze the unsymmetric relationship between events. As the thunder and rain example introduced in Section 1 the unsymmetric commonly exist in the real-world application.

G2: strength expansion from spatial temporal view

3.2 Task

(T1): Calculate the correlation between two climate events. Understanding the relationships between different climate phenomena is critical to understanding climate patterns, predicting future trends, and I dentifying potential impacts on environmental change and human activities. The task of computing a correlation between two climate events aims to quantify the strength of their relationship, providing climate scientists with valuable information.

(T2): Allows comparison of correlations across time or space. When analyzing, we don't just want to know whether the climate change will be caused by different locations, it can help us understand how much the difference in the environment affects. After a long time, whether the development of human society also has an impact. These are the reasons for the impact, so it should support the comparison of correlations in different times or spaces to achieve a more complete analysis.

(T3): Design an interactive visual interface. One of the main tasks in climate analysis is to design interactive visualization interfaces that enable users to explore and understand complex climate data. The development of effective tools is very helpful for experts to obtain detailed information about their interests, which also allows experts to better interpret and verify the calculated results

4 METHOD

4.1 Dataset

The dataset used in this study is REACHES (Reconstructed East Asian Climate History Encoding System), which was created by collating climate-related texts from ancient books and documents of the Ming and Qing Dynasties in China. The time range of the dataset is from 1368 to 1911, and the spatial scope covers the whole of China. It is a data set with a very large time span and spatial span. The dataset contains 32 attributes, such as the start and end time, location, and source of each event. The most important attribute is the "event code", which converts each climate data into a nine-digit code. All the climate records in ancient books are sorted and classified into several major categories such as "precipitation", "temperature" and "thunder and lightning", and some major categories are divided into subcategories according to the different formation methods. For example, the "Precipitation" category is divided into "Rain", "Snow", and "Hail". In addition to climate-related categories, the dataset also records climate-induced disaster events, which allows us to not only study the correlation between climates but also look for the relationship between climate events and disasters. Since this dataset is collected from ancient books and documents, there are some unavoidable problems. The closer to the modern data, the advanced nature of technological development, or the shorter time interval makes it have a more complete record. In addition, places close to cities or coastal areas will have more records due to the dense population. Or the record at that time itself is not clear. These problems lead to data imbalance and missing data values in this data. Even so, many climatologists use this dataset for research.

4.2 Data preprocessing

Algorithm 1 Events' intersection

for all X **do**
  if $Y_{\text{year}} = X_{\text{year}}$ and $Y_{\text{distance}} - X_{\text{distance}} \leq 100\,\text{km}$ **then**
    X is intersection Y
  end if
end for

In the part of data pre-processing, experts gave us many suggestions based on their domain knowledge. The first is to select the attributes to use. Although the REACHES dataset is very detailed in terms of attributes because it is historical text data, there are many missing values for the event's time attributes, and most of the time attributes except the start year are missing in records, so we only consider the year in which events occurred when calculating correlations between events. The second step is to define whether there is an intersection between two events. In order to calculate frequent itemsets between events, we must first understand which events occur together, that is, which events have intersections. The calculation method is shown in Algorithm(1), we use the judgment criteria suggested by experts. For each type of X event in every record, look for Y events that occurred in the same year as X, and their distance is less than 100 kilometers. If there is a Y that meets the conditions, they will be regarded as intersections, and prepared for using the Apriori algorithm later.

4.3 Apriori

The Apriori algorithm(2) is a widely used frequent pattern mining algorithm. Its purpose is to mine the correlation between item sets from a large amount of data and generate valuable association rules. The algorithm works as follows:

1. The first step is to find all frequent itemsets of a single item. This means calculating the support (eq.1) (occurrence frequency) of each item in the entire dataset and then selecting items with support greater than a preset threshold (called minimum support) as frequent itemsets.

2. Then generate candidate k-itemsets based on frequent (k-1) itemsets. This step is called the apriori-gen operation. Specifically, for each itemset in the frequent (k-1) itemsets, compare their top k-2 items, and if the top k-2 items are the same, merge the two itemsets to generate the candidate k items set.

3. Every time step 2 is executed, for each candidate itemset, check whether its support is greater than or equal to the minimum support threshold. If so, add the itemset to the frequent itemset.

4. Finally, all retained frequent itemsets compose all item association rules in the dataset.

$$Support(X,Y) = \frac{freq(X,Y)}{N} \quad (1)$$

After the frequent itemsets are found, the confidence can be calculated. Confidence is used to measure the credibility of an association rule(eq.2). Taking XY itemset as an example, there will be two situations of $X \rightarrow Y$ and $Y \rightarrow X$, where $X \rightarrow Y$ means that when event X occurs, the probability of event Y occurring, that is, the conditional probability $P(B|A)$. And setting the minimum confidence can help us filter out association rules with too low

Algorithm 2 Apriori Algorithm

$L1 = \{\text{large 1-itemsets}\}$
for $k = 2; L_{k-1} \neq \phi; k \mathrel{+}= 1$ **do**
  $C_k = \text{apriori-gen}(L_{k-1})$ ▷ Generate candidate itemsets
  for all $t \in D$ **do**
    $C_t = \text{subset}(C_k, t)$
    for all $c \in C_t$ **do**
      $c.\text{count} \mathrel{+}= 1$
    end for
  end for
  $L_k = \{c \in C_k \mid c.\text{count} \geq \text{minsup}\}$
end for
$\text{Answer} = \bigcup_k L_k$ ▷ Final set of frequent itemsets

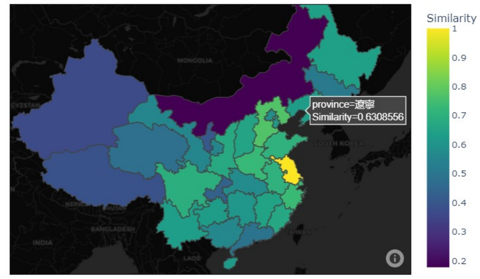

Figure 2: map

confidence. Only association rules with confidence scores greater than or equal to the threshold are considered ”credible”.

$$Confidence(X \rightarrow Y) = \frac{P(X \cap Y)}{P(X)} \tag{2}$$

5 Visual Design

In this section, we introduce our interactive visualization system, which includes different graphs to help users explore the association rules in the REACHES dataset.

5.1 Map

The display of the map can help users understand the relationship between geographical space, and by marking the location on the map, they can quickly understand the distance in different locations and their relative locations. This helps us better understand the connections between neighboring places. We built a zoomable map (Fig.2) of China where we use color encoding to show the differences between provinces. According to the user’s operations on different charts, the color of the provinces on the map will also change. For example, when the user compares the arc diagram graph of different provinces, the color of the provinces will change according to their similarity. When comparing the differences in the association of specific events in different time spaces, the differences in different provinces will also be displayed according to the dynasty clicked by the user. Using maps and other visual charts, we can quickly observe and analyze the overall trends and patterns of the data, which helps users extract valuable information.

5.2 Association rule graph

In the association rule graph (Fig.3), our visualization system will calculate the association rules according to the provinces selected by the user, and then filter the calculated results by the threshold of the minimum confidence value. Here, users are allowed to change the threshold to define what is strong association rules and draw the association rules that meet the conditions into a graph. In order to make it easier to find out whether there are differences or similar rules between the two provinces, we use an arc diagram graph to represent them. Because of its simplicity and readability, users can quickly understand the composition of association rules in each province or the differences between different provinces. Through the results of the association rule graph, the user can determine an association rule for more detailed analysis.

5.3 Line chart

In this figure, we use a line chart to show the similarity between different provinces across all dynasties, where each line represents a province. After the user selects the base province, in addition to updating the color of the province on the map, a line chart will also be generated to help the user explore the differences between the provinces in different dynasties. In the part of generating the line graph, we first calculate the frequent item sets of the two provinces through the Apriori algorithm and sort out all the same combinations, and then use the Euclidean distance to calculate the confidence value of all common combinations as the similarity. Through this line chart, users can clearly see the difference between other provinces and the base province under the same dynasty, and it can also allow users to discover whether there is a province under a certain dynasty, its similarity with Other dynasties are markedly different.

5.4 Heatmap

In our visualization tool, we use a total of two sets of heatmaps. The first set is used to compare the changes of an original correlation if the third event is added. The event on each row represents the original event associated with the selected event, and the event in each column is the third type of event to be considered additionally. Through the Apriori algorithm, we can find out the original combination pattern and the pattern adding the third event, and calculate the difference in confidence value between these two situations to represent the degree of influence of adding the third event. Users can select the events they are curious about through the arc diagram graph, and the two heatmaps can compare the correlation of these event combinations in specific spatiotemporal patterns and all spatiotemporal patterns.

In the second set of heatmaps, we calculate the confidence values in all provinces and all dynasties based on the two specific events that the user wants to study. Users can use this heatmap to see whether there is a certain combination of time and space that has a confidence value significantly different from other combinations. In addition, when the user selects a dynasty, the confidence values of all provinces in this dynasty will also be updated on the map.

6 User Study

In this section, we introduce two use cases to explain what experts have discovered through our visualization system. To prove that our visualization system can achieve the following two goals: (1) Find out the correlation of events under a specific combination of time and space. (2) ......

6.1 Pestilence’s cause

In this example, we first use the line chart to compare the difference between a single province and the whole of China in different dynasties and find that during the ”Qianlong” and ”Jiaqing” periods, the association rule composition of each province in northern provinces and the association rule of all provinces in China have noticeable differences. Therefore, we choose to use the arc diagram to display the complete association rule between these provinces. By adjusting the confidence values, it can be found that the provinces in the northern region show that the event ”Pestilence” is more often correlated with other events, especially the two correlations *PestilenceandFamine*

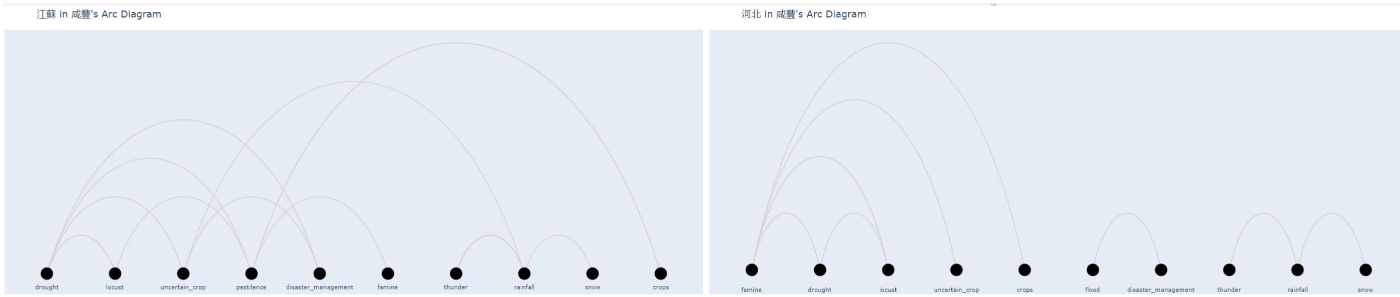

Figure 3: Caption

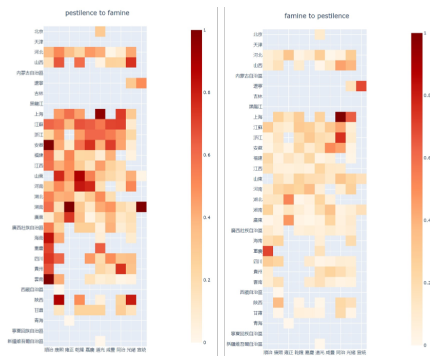

Figure 4: Pestilence

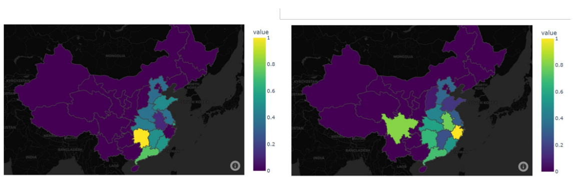

Figure 5: Pestilence's main cause in "YongZheng". Left is $Pestilence \rightarrow Famine$. Right is $Pestilence \rightarrow Flood$.

and $PestilenceandFlood$. Interestingly, this result was not observed for the entire range's association rule.

In order to analyze this phenomenon more deeply, we displayed the confidence values of these two associations on a heatmap across all provinces and dynasties. Since the confidence value is calculated using conditional probabilities, the order of events is meaningful, and reversing the events can lead to different results. We find that the rule: $Pestilence \rightarrow Famine$ had confidence values above 0.5 for nearly one-third of the provinces in each dynasty. However, the rule: $Famine \rightarrow Pestilence$ consistently had lower confidence values for all combinations (Fig.4). Similarly, in the rule: $PestilenceandFlood$, we observed similar results. From this, we can conclude that "Famine" or "Flood" is one of the primary causes of the occurrence of the "Pestilence" disaster, but "Pestilence" is not a primary cause for the occurrence of "Famine" or "Flood".

Finally, we match the confidence value displayed on the map for every dynasty, and we can find that in the same dynasty, the causes of "Pestilence" in different provinces are different(Fig.5). Because of this result and the lack of historical records of "Pestilence" in some places, it is very difficult to find the connection between "Pestilence" and other events if we want to find the connection from the correlation combination of the whole China.

6.2

7 Conclusion

Climate data can provide valuable information. Understanding and identifying the correlations between climates is a crucial task in climate analysis, particularly when studying ancient climates, which is a topic of great interest to many meteorologists. However, data on ancient climates can only be collected from historical records or literature, which makes studying ancient climates more difficult. In this study, we use the REACHES (Reconstructed East Asian Climate Historical Encoded Series) dataset. We designed an approach based on the FP-Growth algorithm to discover frequently occurring associations among climate events. Based on these associations, we developed an interactive visualization tool that allows experts to explore the historical climate of the Ming and Qing dynasties spanning 600 years. This tool helps them understand the interactions between climates, enabling them to investigate the relationships between events and undertake more advanced research tasks, such as reconstructing historical climates or comparing the impacts among different climates.

References

[1] R. Agrawal, R. Srikant, et al. Fast algorithms for mining association rules. In *Proc. 20th int. conf. very large data bases, VLDB*, vol. 1215, pp. 487–499. Santiago, Chile, 1994.

[2] T. Anwar, C. Liu, H. L. Vu, M. S. Islam, and T. Sellis. Capturing the spatiotemporal evolution in road traffic networks. *IEEE Transactions on Knowledge and Data Engineering*, 30(8):1426–1439, 2018. doi: 10.1109/TKDE.2018.2795001

[3] A. G. Asuero, A. Sayago, and A. González. The correlation coefficient: An overview. *Critical reviews in analytical chemistry*, 36(1):41–59, 2006.

[4] N. Aubry, R. Guyonnet, and R. Lima. Spatiotemporal analysis of complex signals: theory and applications. *Journal of Statistical Physics*, 64:683–739, 1991.

[5] R. Brázdil, P. Dobrovolnỳ, M. Trnka, L. Řezníčková, L. Dolák, and O. Kotyza. Extreme droughts and human responses to them: the czech lands in the pre-instrumental period. *Climate of the Past*, 15(1):1–24, 2019.

[6] R. Brázdil, A. Kiss, J. Luterbacher, D. J. Nash, and L. Řezníčková. Documentary data and the study of past droughts: a global state of the art. *Climate of the Past*, 14(12):1915–1960, 2018.

[7] S. Brönnimann, O. Martius, C. Rohr, D. N. Bresch, and K.-H. E. Lin. Historical weather data for climate risk assessment. *Annals of the New York Academy of Sciences*, 1436(1):121–137, 2019.

[8] H. J. Bruins. Ancient desert agriculture in the negev and climate-zone boundary changes during average, wet and drought years. *Journal of Arid Environments*, 86:28–42, 2012.

[9] A. J. Challinor, J. Watson, D. B. Lobell, S. Howden, D. Smith, and N. Chhetri. A meta-analysis of crop yield under climate change and adaptation. *Nature climate change*, 4(4):287–291, 2014.

[10] A. Hall and G. V. Jones. Spatial analysis of climate in winegrape-growing regions in australia. *Australian Journal of Grape and Wine Research*, 16(3):389–404, 2010.

[11] J. Han, J. Pei, and Y. Yin. Mining frequent patterns without candidate generation. *ACM sigmod record*, 29(2):1–12, 2000.

[12] J. Hansen, M. Sato, R. Ruedy, P. Kharecha, A. Lacis, R. Miller, L. Nazarenko, K. Lo, G. Schmidt, G. Russell, et al. Climate simulations for 1880–2003 with giss modele. *Climate dynamics*, 29:661–696, 2007.

[13] Z. Hao, J. Zheng, X. Zhang, H. Liu, M. Li, and Q. Ge. Spatial patterns of precipitation anomalies in eastern china during centennial cold and warm periods of the past 2000 years. *International Journal of Climatology*, 36(1):467–475, 2016.

[14] K.-H. E. Lin, P. K. Wang, P.-L. Pai, Y.-S. Lin, and C.-W. Wang. Historical droughts in the qing dynasty (1644–1911) of china. *Climate of the Past*, 16(3):911–931, 2020. doi: 10.5194/cp-16-911-2020

[15] K.-b. Liu, C. Shen, and K.-s. Louie. A 1,000-year history of typhoon landfalls in guangdong, southern china, reconstructed from chinese historical documentary records. *Annals of the Association of American Geographers*, 91(3):453–464, 2001.

[16] A. N. Mackintosh, B. M. Anderson, and R. T. Pierrehumbert. Reconstructing climate from glaciers. *Annual Review of Earth and Planetary Sciences*, 45:649–680, 2017.

[17] S. T. Mai, H. T. Phi, A. Abubakar, P. Kilpatrick, H. Q. V. Nguyen, and H. Vandierendonck. Dengue fever: From extreme climates to outbreak prediction. In *2022 IEEE International Conference on Data Mining (ICDM)*, pp. 1083–1088, 2022. doi: 10.1109/ICDM54844.2022.00135

[18] M. A. Moritz. Spatiotemporal analysis of controls on shrubland fire regimes: age dependency and fire hazard. *Ecology*, 84(2):351–361, 2003.

[19] Q. Pei, H. F. Lee, D. D. Zhang, and J. Fei. Climate change, state capacity and nomad–agriculturalist conflicts in chinese history. *Quaternary International*, 508:36–42, 2019.

[20] G. L. Schaefer, M. H. Cosh, and T. J. Jackson. The usda natural resources conservation service soil climate analysis network (scan). *Journal of Atmospheric and Oceanic Technology*, 24(12):2073–2077, 2007.

[21] F. H. Schweingruber and K. R. Briffa. Tree-ring density networks for climate reconstruction. In *Climatic variations and forcing mechanisms of the last 2000 years*, pp. 43–66. Springer, 1996.

[22] J. Slingo and T. Palmer. Uncertainty in weather and climate prediction. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 369(1956):4751–4767, 2011.

[23] W. Song and X. Ma. Research on climate prediction based on deep learning. In *2021 3rd International Symposium on Smart and Healthy Cities (ISHC)*, pp. 107–109, 2021. doi: 10.1109/ISHC54333.2021.00028

[24] P. Tan and H. Liao. Reconstruction of temperature, precipitation and weather characteristics over the yangtze river delta area in ming dynasty. *J. Geogr. Sci*, 57:61–87, 2012.

[25] R. Taylor. Interpretation of the correlation coefficient: a basic review. *Journal of diagnostic medical sonography*, 6(1):35–39, 1990.

[26] H. Tian, L. C. Stige, B. Cazelles, K. L. Kausrud, R. Svarverud, N. C. Stenseth, and Z. Zhang. Reconstruction of a 1,910-y-long locust series reveals consistent associations with climate fluctuations in china. *Proceedings of the National Academy of Sciences*, 108(35):14521–14526, 2011.

[27] R. Vassoler and G. Zebende. Dcca cross-correlation coefficient apply in time series of air temperature and air relative humidity. *Physica A: Statistical Mechanics and its Applications*, 391(7):2438–2443, 2012.

[28] G.-J. Wang, C. Xie, S. Chen, J.-J. Yang, and M.-Y. Yang. Random matrix theory analysis of cross-correlations in the us stock market: Evidence from pearson's correlation coefficient and detrended cross-correlation coefficient. *Physica A: statistical mechanics and its applications*, 392(17):3715–3730, 2013.

[29] H. Wang, Y. Lu, S. T. Shutters, M. Steptoe, F. Wang, S. Landis, and R. Maciejewski. A visual analytics framework for spatiotemporal trade network analysis. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):331–341, 2019. doi: 10.1109/TVCG.2018.2864844

[30] M. Watanabe, T. Suzuki, R. O'ishi, Y. Komuro, S. Watanabe, S. Emori, T. Takemura, M. Chikira, T. Ogura, M. Sekiguchi, et al. Improved climate simulation by miroc5: mean states, variability, and climate sensitivity. *Journal of Climate*, 23(23):6312–6335, 2010.