PATTERN EXTRACTION BY USING BOTH SPATIAL AND TEMPORAL FEATURES ON TURKISH METEOROLOGICAL DATA

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

IŞIL GOLER

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
COMPUTER ENGINEERING

DECEMBER 2010

Approval of the thesis:

PATTERN EXTRACTION BY USING BOTH SPATIAL AND TEMPORAL FEATURES ON TURKISH METEOROLOGICAL DATA

submitted by **IŞIL GOLER** in partial fulfillment of the requirements for the degree of **Master in Computer Engineering Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen
Dean, Graduate School of **Natural and Applied Sciences** _______________

Prof. Dr. Adnan Yazıcı
Head of Department, **Computer Engineering** _______________

Prof. Dr. Adnan Yazıcı
Supervisor, **Computer Engineering Dept., METU** _______________

Asst. Prof. Dr. Pınar Şenkul
Co-Supervisor, **Computer Engineering Dept., METU** _______________

Examining Committee Members:

Prof. Dr. Hakkı Toroslu
Computer Engineering Dept., METU _______________

Prof. Dr. Adnan Yazıcı
Computer Engineering Dept., METU _______________

Assoc. Prof. Zuhal Akyürek
Geodetic and Geographic Information Technologies Dept., METU _______________

Asst. Prof. Dr. Pınar Şenkul
Computer Engineering Dept., METU _______________

Asst. Prof. Dr. Tolga Can
Computer Engineering Dept., METU _______________

Date: 27.12.2010

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: IŞIL GOLER

Signature :

ABSTRACT

PATTERN EXTRACTION BY USING BOTH SPATIAL AND TEMPORAL FEATURES ON TURKISH METEOROLOGICAL DATA

Goler, Işıl
M.Sc., Department of Computer Engineering
Supervisor: Prof. Dr. Adnan Yazıcı
Co-Supervisor: Asst. Prof. Dr. Pınar Şenkul

December 2010, 100 pages

With the growth in the size of datasets, data mining has been an important research topic and is receiving substantial interest from both academia and industry for many years. Especially, spatio-temporal data mining, mining knowledge from large amounts of spatio-temporal data, is a highly demanding field because huge amounts of spatio-temporal data are collected in various applications. Therefore, spatio-temporal data mining requires the development of novel data mining algorithms and computational techniques for a successful analysis of large spatio-temporal databases. In this thesis, a spatio-temporal mining technique is proposed and applied on Turkish meteorological data which has been collected from various weather stations in Turkey. This study also includes an analysis and interpretation of spatio-temporal rules generated for Turkish Meteorological data set. We introduce a second level mining technique which is used to define general trends of the patterns according to the spatial changes. Genarated patterns are investigated under different temporal sets in order to monitor the changes of the events with respect to temporal changes.

Keywords: Spatio-Temporal Data Mining, Mining Rules, Spatio-Temporal data

ÖZ

TURKİYE METEOROLOJİ VERİSİNİN UZAYSAL VE ZAMANSAL ÖZELLİKLERİNİ KULLANARAK KURALLAR BULUNMASI

Goler, Işıl
Yüksek Lisans, Bilgisayar Mühendisliği Bölümü
Tez Yöneticisi: Prof. Dr. Adnan Yazıcı
Ortak Tez Yöneticisi: Dr. Pınar Şenkul

Aralık 2010, 100 sayfa

Bilgi yığınlarının boyutunun artmasıyla veri madenciliği önemli bir araştırma konusu haline geldi ve hem endüstride hem de akademik dünyada büyük ilgi gördü. Uzaysal ve zamansal veri madenciliği, büyük miktarda uzaysal ve zamansal veri içeren bilgi yığınlarından önemli bilgi veya desen çıkarılması olup, çok çeşitli uygulamalarda uzaysal ve zamansal veri toplandığından gittikçe ilgi çeken bir konu haline geldi. Uzaysal ve zamansal veri madenciliğinde başarılı bir analiz için yeni algoritmalara ve hesaplama yöntemlerine ihtiyaç vardır. Bu tezde, yeni bir yaklaşım geliştirilerek, bu yaklaşım Türkiye'deki farklı hava istasyonlarından alınmış meteoroloji verisi üzerinde uygulandı. Ayrıca, bu veri kullanılarak çıkarılmış desenler analiz edilerek yorumlandı. Ortaya koyduğumuz ikinci seviye veri madenciliği tekniği ile çıkarılan desenlerin uzaysal değişikliklere göre genel trendleri çıkarıldı. Benzer bir teknikle, çıkarılan desenler farklı zamansal dönemler için incelenerek, zaman değişiminin desenler üzerindeki etkisi analiz edildi.

Anahtar Kelimeler: Uzaysal ve zamansal veri, uzaysal ve zamansal veri madenciliği

To My Grandparents

This thesis is dedicated to my lovely and compassionate grandparents who recently passed away. May they rest in peace.

ACKNOWLEDGEMENTS

I particularly thank my supervisor, Prof. Dr. Adnan Yazıcı, whose encouragement, guidance and support from the initial to the final level enabled me to complete this thesis. Thanks for the enthusiasm and inspiration which was always there when I needed it.

I am heartily thankful to my co-supervisor Dr. Pınar Şenkul for her valuable supervision, suggestions and comments during my thesis.

Also I thank Havelsan Inc. for providing an opportunity to study for the degree of master.

I can not finish without saying how grateful I am to my brother Güneş Goler who is the one who convinced me to study at METU and supported me in any respect during the completion of my education.

Finally, and the most importantly, I wish to thank my parents for their complimentary love and support. Thank you for having me. Now look what I have done!

.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CENG	Computer Engineering
IEEE	Institute of Electrical and Electronic Engineers
METU	Middle East Technical University
GIS	Geographic Information System

CHAPTER 1

INTRODUCTION

1.1 Motivation and Scope

Spatio-temporal data mining is an emerging research area that is dedicated to the development of novel algorithms and computational techniques for the successful analysis of large spatio-temporal databases. Huge collections of spatio-temporal data often contain possibly interesting information and valuable knowledge. It is obvious that a manual analysis of these data is impossible and data mining should provide useful tools and technology in this setting. Necessity of techniques for the analysis of huge collections of spatio-temporal data is the motivation of this study.

Starting point of this study is the future directions stated in [1]. The same meteorological data as in [1] is used for finding spatio-temporal patterns. In contrast to earlier research on geospatial data mining, which take a static view of data (e.g. location, dimensionality, geometry, and topology) in order to capture the spatiality [2], we aimed to integrate the temporal dimension into the analysis of geospatial data since all geographic phenomena evolve in time. Therefore, the main focus of this study is to propose an approach that takes into consideration both spatial and temporal aspects of the meteorology data set. After literature survey, the approach given in [9], which considers spatial and temporal data together, has been extended in several directions as presented in the following subsections. The most important extension of the approach is the second level mining technique which is used for finding the general trends of the generated patterns. Also, with a similar technique, generated patterns are investigated under different temporal sets in order to monitor the changes of the events with respect to temporal changes.

1.2 Related Work

Although data mining is a heavily studied research subject, the number of studies on spatio-temporal mining techniques is comparatively limited. In the rest of this section, an overview of studies on spatio-temporal data mining is presented.

For a general panorama of the data mining field, [3] is a good reference to commence the literature survey. It provides the basic information about the history of data mining applications, which helps to select possible directions for research.

The work presented in [4] explains the importance of considering both spatiality and temporality of the geographic data for a better understanding of geographic process and events. This study states that "*A crucial challenge in spatio-temporal data mining is the exploration of efficient methods due to the large amount of spatio-temporal data and the complexity of spatio-temporal data types, data representation, and spatial data structure*"[4]. According to [4], data mining techniques should be modified so that they can exploit the rich spatial and temporal relationships/patterns embedded in the datasets and it presents a classification of spatio-temporal data mining tasks and techniques. It groups basic tasks in five groups as segmentation, dependency analysis, deviation and outlier analysis, trend discovery and generalization and characterization. Segmentation is described as clustering and classification. Dependency analysis is about finding rules to predict the value of some attribute based on the value of other attributes over time. Outlier analysis is for finding data items that exhibit unusual deviations from expectations. Trend discovery is used for discovering correlations among the events in sequences. Characterization is the compact description of the data. The techniques employed for each of these tasks are listed. This paper indicates the absence of efficient spatio-temporal data mining techniques and points the necessity of an effective technique.

In [5], the authors investigate the discovery of spatial association rules in geographic information databases [5]. Although the temporal aspect of the data is not considered in their study, it is a good starting point for understanding spatial association rule concept. Another important aspect is that the technique has been applied on geographic data. The researchers have proposed an efficient method for mining strong spatial association rules in geographic information databases. The proposed technique firstly searches at a higher concept level for large patterns and strong implication relationships among the large patterns are searched at a coarse resolution scale. Then only for those large patterns, it deepens the search to lower concept levels. Such a deepening search process continues until no large patterns can be found. With this approach they are able to reduce the amount of computation. Also many other relevant studies are analyzed throughout the paper. Spatial association rules and relationships which include both spatial and non-spatial predicates are given in detail by examples. Concept hierarchies are provided for both spatial and non-spatial predicates to facilitate mining multiple-level association rules and efficient processing. Multiple-level association rule concept defined in the paper is to find the rules in a higher hierarchy and searching for the frequent rules in the lower hierarchies. For example if a frequent rule is found for water then search is deepened for rivers and lakes which are the subsets of water. Although multiple-level association rule concept defined in this paper is significantly different from the concept defined by this thesis, "Mining Rules" modality for mining generated rules is an outcome of the key word of 'multiple-level rules'.

[6] demonstrates an application of association rule mining to spatio-temporal data with a case study. This case study concerning urban growth in the Denver region demonstrates how the association rule mining may be applied to spatio-temporal data. The development of concept hierarchies through data classification shows a methodology that supports multiple level spatio-temporal association rule mining and thereby explores the effect of attribute resolution on the generation of interesting rules. Examining this case study provides a valuable opportunity to observe the result of methodologies when they are applied on real data sets.

[7] concerns the temporal aspect of the spatio-temporal data. Temporal co-orientation pattern mining which is the problem of temporal aspects of spatial co-orientation patterns is introduced. Temporal co-orientation patterns represent how spatial co-orientation patterns change over time. Researches propose the three-stage approach which transforms the problem into sequential pattern mining. This paper points other interesting research studies which are based on the neighboring relationships for the spatial aspect.

Fuzziness might have been used for the temporal aspect of the approach defined by this thesis. In this regard, [8] is analyzed. [8] presents an algorithm for mining fuzzy temporal patterns from a given process instance. Fuzzy representation of time intervals embedded between the activities is used for that purpose. Weighted temporal graphs are used to represent temporal relationships. Weighted temporal graph is a term which is introduced in this paper. According to the definition, the weight of an arc is the time between the corresponding activities. Processes are sorted on the graphs due to their starting and ending times. Overlapping and following events can be followed from the produced graphs. As a result of these graphs, frequent patterns are found such as “Having activity A is finished, activity B starts in a short period and then activity C starts in a long period”. Temporal relationships in short, medium, long periods represent the fuzzy instances. And the difference of this study is in its approach which takes into consideration the temporal relationships between activities in order to have a detailed knowledge of the process flow.

Because of the urgent need to discover interesting time and space patterns in huge spatio-temporal databases, a new concept named “flow patterns” is introduced. J. Wang, et al. [9] present a disk-based algorithm, *FlowMiner*, which utilizes temporal relationships and spatial relationships amid events to generate flow patterns. According to the study Flow patterns are intended to describe the change of events over space and time. The aim of the study is discovering computationally challenging spatio-temporal patterns such as; ‘the hurricane season typically begins in late October through early February in the South-Western part of the United States, and the peak in the sales of the North America region is frequently followed by a peak in sales in Asia within a 2-month period”. Another exciting point in the study is the contribution which says that the study is the first work that takes into consideration both spatial and temporal information in the mining of spatio-temporal patterns. Neighboring relationships which define the spatial associations and temporal time window definitions which establish the temporal relationships are well defined and they are suitable to be used on our dataset. Details of the *FlowMiner* algorithm and the definitions of the terms of this study are given in detail in the background section.

[10] is an improvement and generalization of [9]. [10] introduces a new class of spatio-temporal patterns, called generalized spatio-temporal patterns, to describe the repeated sequences of events that occur within small neighborhoods. An algorithm named *GenSTMiner* based on the idea of pattern growth approach is developed. The difference between this study and the study by J. Wang, et al. [9] is that while flow patterns can clearly capture the flow of events to some degree, J. Wang, et al. in [9] relies heavily on the assumption that these events will repeat themselves in exactly the same locations. However, in some applications, it is observed that the absolute locations in which event *e* has occurred are not important. Rather, the relative locations of events with respect to the event *e* are interesting. Generalized spatio-temporal patterns are defined to summarize these sequential relationships between events that are prevalent in sharing the same topological structures. Difference of flow patterns when they are compared to generalized spatio-temporal patterns can be clarified with the following example; Flow patterns are able to capture the flow of events such as; an increase of air temperature at location *x* leads to an increase in wind speed and gust speed at location *y*; and an increase of air temperature at location *z* leads to an increase in wind speed and gust speed at location *t*. Flow patterns are unable to provide a general trend. On the other hand, the generalized spatio-temporal pattern reveals the trend for the given example that whenever there is an increase of air temperature at a specific location, an increase in wind speed and gust speed is expected at its northeast neighbor. And by knowing the general trend, the meteorologist is able to perform more accurate forecast of the weather.

Although the idea of the generalizing the found patterns results in more general and useful trends, it is hard to apply this idea on the available data set. There are many locations that the same kinds of events occur, and it is hard to choose the reference locations to generalize the eventsets. “Mining Rules’ phase of the approach defined by this thesis is implemented for a similar purpose (for finding general trends). Details of the *GenSTMiner* algorithm and the definitions of the terms of this study are given in detail in the background section.

Because of the need of an efficient mining algorithm that can discover complex relationships among events with duration, an algorithm called *IEMiner* is designed by [11]. And to the best of author’s knowledge [11] is the first work to build an interval-based classifier. Until this study, existing temporal mining techniques used to assume that events do not have any duration. However, events in many real world applications have durations, and the relationships among these events are often complex. Some of the sequential patterns are inadequate to express the complex temporal relationships in domains such as meteorology, multimedia and finance. By the implementation defined in this study, these complex temporal relations are clarified by the methodology defined in the paper. Events are related to each other due to their start and end times, overlapping and meeting situations. A hierarchical representation is designed for holding the contain count, finish by count, meet count, overlap count and start count to differentiate all the possible cases. This augmented hierarchical representation is lossless. Apriori-based *IEMiner* algorithm to mine frequent temporal patterns from interval-based events might have been used in the approach defined by this thesis as well; however time window definitions of the previous studies are used with some modifications instead.

1.3 Contributions

Although the approach proposed by this study is inspired by J. Wang, et al. [9], it is extended in various aspects as follows:

- In order to apply the approach in [9] on the available data set, preprocessing phase, which does not exist in [9], is implemented. Meteorology data set is stored in different database tables, events are determined and they are classified due to their intensities.

- Used data structures and processing of these structures in this study are different from that of [9]. In [9] a tree structure is used for the candidate generation phase. In our work, we used eventsets. Eventsets which include the events and their locations are compared with other eventsets due to the neighborhood and time relationships. Flow patterns which are composed of these eventsets are stored in the nodes of the tree. Eventsets in the nodes are extended with the other eventsets only if they are frequent. In the approach defined by this thesis, an easier way is selected for the candidate generation phase. Weather stations are selected one by one if any events have occurred in them. Consecutive eventsets that includes the events of the selected station and events of its neighbor stations are generated. All these consecutive eventsets are stored in database. Frequencies of the eventsets are considered during the mining phase. Generation of all possible relations and storing them in database tables which causes many database transactions to reach the stored data during mining phase is not so efficient when the performance of the implementation is considered, but simplicity of the technique and easiness of its implementation has an important advantage. As a result, all the phases are conducted by different techniques which are specific to this study.

- Instead of using strict temporal intervals as in [9] for defining the temporal relationships, sliding intervals are implemented to avoid losing the temporal relations of two different intervals. Because when the strict intervals are used, only the relations in the same time intervals are considered. But by the sliding mechanism, intervals are built dynamically by including all of the possible correlations. It is one of the most important extensions of the inspired approach.

- A basic example has been illustrated in [9] with a small data set. Applying a similar idea on a real data set which is really huge and getting meaningful results is an important accomplishment of this thesis work.

- Another important aspect that differs from the approach defined in [9] is the addition of the "Mining Rules" phase which is proposed for the generalization of the rules. All the frequent patterns are collected, they are categorized due to their event types, and by a second pass over these categorized frequent patterns we find out the general trend of the events according to the location changes. And to the best of our knowledge this is the first work that takes into consideration the second level mining by the proposed technique.

- Event changes with respect to temporal changes are analyzed. These experiments demonstrate how the proposed approch can be used for many different and useful investigations.

1.4 Organization of the Thesis

The preceding sections of this chapter introduce the motivation and scope of the study and present the summary of the thesis application. The content of the remaining chapters are as follows: Chapter 2 provides literature survey and background information, Chapter 3 explains the proposed approach, Chapter 4 describes the implementation details with sample source codes and class views, Chapter 5 presents the interpretation of the output files which contains the generated spatio-temporal patterns. This chapter also includes the evaluation of the results and experiments applied on the generated patterns.

CHAPTER 2

BACKGROUND

In this chapter, the background information is summarized from the literature which guides the reader to track the details clearly in the subsequent chapters.

2.1 Data Mining

Data mining is the extraction of interesting non-trivial, implicit, previously unknown and potentially useful information or patterns from data in large databases. Data mining involves the use of sophisticated data analysis tools to discover previously unknown, valid patterns and relationship in large data sets.

Data mining can be performed on data represented in quantitative, textual, or multimedia forms. Data mining applications can use a variety of parameters to examine the data. They include [12]

- association (patterns where one event is connected to another event, such as purchasing a pen and purchasing paper),
- sequence or path analysis (patterns where one event leads to another event, such as the birth of a child and purchasing diapers),
- classification (identification of new patterns, such as coincidences between duct tape purchases and plastic sheeting purchases),
- clustering (finding and visually documenting groups of previously unknown facts, such as geographic location and brand preferences),
- forecasting (discovering patterns from which one can make reasonable predictions regarding future activities, such as the prediction that people who join an athletic club may take exercise classes).

The Data Mining process usually consists of three phases [13]: 1) pre-processing or data preparation, 2) modeling and validation and, 3) post-processing or deployment. During the first phase, the data may need cleaning and transformation according to possible constraints imposed by some tools, algorithms, or users. The second phase consists of choosing or building a model that better reflects the application behavior. Such a model should be evaluated in terms of its efficiency and accuracy of its predictive results. Finally, the third step consists of using the selected model to effectively study the application behavior.

Usually, the model output requires some post-processing in order to exploit it. This step can benefit from data visualization, since interactivity and user expertise are very important in the final decision-making and data interpretation.

2.2 Spatio-Temporal Data Mining

Spatial data mining is an emerging research area dedicated to the development and application of novel, typically inductive, computational techniques for the analysis of very large, heterogeneous spatial databases [14]. Spatio-temporal data mining takes into consideration the dynamics of spatially extended systems for which large amounts of data exist. Given that all real world spatial data exists in some temporal context, and knowledge of this context is often essential in interpreting it, spatial data mining is inherently spatio-temporal data mining to some degree. Spatio-temporal data mining is used in many application areas such as GIS, robotics, computer vision, computational biology, mobile computing and traffic analysis.

Spatio-temporal data mining presents a number of challenges due to the complexity of geographic domains, the mapping of all data values into a spatial and temporal framework, and the spatial and temporal autocorrelation exhibited in most spatio-temporal data sets [15].

2.3 Flow Patterns

Flow patterns aim to link event changes in one location to another location in order to reveal insights that can not be obtained otherwise. Discovering flow patterns is a challenge because of the potentially large search space and the large number of candidates. Concept of flow patterns are introduced by incorporating spatial neighborhood relations into sequence pattern mining [9].

Spatio-temporal databases capture both time and space dimensions. First, time is divided into disjoint time windows of length W. Each time window denotes a time period. Time t_1 and t_2 are said to be "near" if they are in the same time period.

Next step is the division of space into the disjoint grid cells. Grid cells are represented as a location set S = $\{l_1, l_2,..l_i\}$. A neighboring relationship R is defined over set S. Location l_1 and l_2 are said to be neighbors if $(l_1,l_2) \in$R. And the *Neighborhood* of a location l is defined as a set of locations N(l) = $\{l_1,..l_k\}$.

A location-based event denoted as $e\ (l,\ t)$ occurring in location l at time t. Two events $e_1(l_1, t_1)$ and $e_2\ (l_2,\ t_2)$, $t_1 <= t_2$, are said to be related iff $(l_1,\ l_2) \in R$ and t_1 is near to t_2. A set of location-based events that occur at the same time is called an *eventset.* An eventset *Ep* at time t_1 is said to *flow* to an eventset *Eq* at time $t_2\ (t_1 <= t_2)$ iff every event in *Ep* is related to every event in *Eq*. And flow of eventsets is denoted as *Ep->Eq*. In addition, an eventset *Et* is said to be *reflexive* iff *Et* flows to itself.

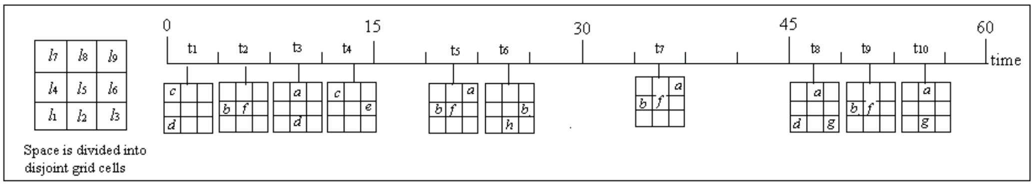


Figure 2-1 Example of a spatio-temporal database (space-time view)

According to the example spatio-temporal database given in Figure 2-1, eventsets can be created as listed in Figure 2-2.

Window ID (wid)	Time	Eventsets
	t_1	$d(l_1)$, $c(l_7)$
	t_2	$b(l_4)$, $f(l_5)$
	t_3	$d(l_2)$, $a(l_8)$
1	t_4	$e(l_6)$, $c(l_7)$
	t_5	$b(l_4)$, $f(l_5)$, $a(l_9)$
2	t_6	$h(l_2)$, $b(l_6)$
3	t_7	$b(l_4)$, $f(l_5)$, $a(l_9)$
	t_8	$d(l_1)$, $g(l_3)$, $a(l_8)$
	t_9	$b(l_4)$, $f(l_5)$
4	t_{10}	$g(l_2)$, $a(l_8)$

Figure 2-2 Data sorted by window id and time

Neighborhood relationship R is set as 1 which means that if the distance of the grid cells is 1, they are said to be neighbors. Two events $d(l_1,\ t_1)$ and $b(l_4,\ t_2)$ are related since they have occured in the same time window, $t_1 <= t_2$ and $(l_1,\ l_4) \in R$ (l_4 is located just above l_1). Eventset $b(l_4)$, $f(l_5)$ at time t_2 *flows* to the eventset $d(l_2)$, $a(l_8)$ at time t_3 since $t_2 <= t_3$ and all the events in the eventsets are related with a neighbouring relationship.

Once the flow patterns are generated, next step is calculating the frequency. A summary tree structure is used to keep track of all the frequent flow patterns that have been generated and to capture the relationships of sequences. Frequent k-flows are placed in the root node. Then they are extended with the frequent flow patterns. *FlowMiner* algortihm which is based on DFS-Scan is used to build the summary tree and to generate the frequent flow patterns.

2.4 Generalized Spatio-Temporal Patterns

Generalized spatio-temporal patterns are generated to summarize the sequential relationships between events that are prevalent in sharing the same topological structures. Notations and definitions of FlowMiner study are also valid for generalized spatio-temporal patterns. But there are some extensions and differences. Locations of the events are represented with their x and y coordinates. Since then, neighboring relationship is calculated in a different way. If R denotes a spatial neighborhood relation over the set of partitioned cells, Two cells (x_1, y_1) and (x_2, y_2) are said to be *neigbors* and it is demonstrated as $<(x_1, y_1), (x_2, y_2)> \in R$, if $|x_1 - x_2| <= n_r$ and $|y_1 - y_2| <= n_r$ where n_r is the number of grid cells. Similarly, the time dimension is divided into disjoint time windows of width W. Two events $e_1(x_1, y_1, t_1)$ and $e_2(x_2, y_2, t_2)$ are said to be *CloseNeighbors* iff $<(x_1, y_1), (x_2, y_2)> \in R$ and $(t_1, t_2) \in W$, denoted as $<e_1(x_1,y_1,t_1), e_2(x_2,y_2,t_2)> \in (R, W)$.

This study is the extension of FlowMiner approach to show the importance of the relative addresses of the events while capturing the invariant topological relationships of a pattern. A reference location is selected in order to incorporate the concept of relative addresses. It is denoted as; $l_{ref} = (x_{ref}, y_{ref})$. And addresses of each occuring event are mapped to their corresponding relative occuring locations as $e_1(x_1 - x_{ref}, y_1 - y_{ref}), e_2(x_2 - x_{ref}, y_2 - y_{ref}), \ldots, e_m(x_m - x_{ref}, y_m - y_{ref})$. A *RelativeEventSet* is a set of mapped events that occur at the same time t. And RelativeEventSets are CloseNeighbor to each other if every event in one of the EventSets is CloseNeighbor of every event in the other EventSet. A *generalized spatio-temporal pattern* is a sequence of RelativeEventsets.

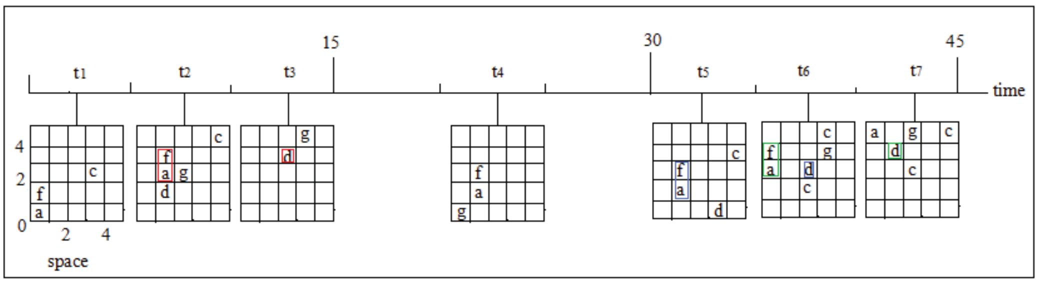


Figure 2-3 Space-time window

Window ID (wid)	Time	Eventsets
1	t_1	a(0,0), c(3,2), f(0,1)
	t_2	a(1,2), c(4,4), d(1,1), f(1,3), g(2,2)
	t_3	d(2,3), g(3,4)
2	t_4	a(1,1), f(1,2), g(0,0)
3	t_5	a(1,1), c(4,3), d(3,0), f(1,2)
	t_6	a(0,2), c(2,1), c(3,4), d(2,2), f(0,3), g(3,3)
	t_7	a(0,4), c(2,2), c(4,4), d(1,3), g(2,4)

Observed Flow Patterns
<a(0,0), f(0,1)> --> d(1,1)
<a(1,2), f(1,3)> --> d(2,3)
<a(1,1), f(1,2)> --> d(2,2)
<a(0,2), f(0,3)> --> d(1,3)

Figure 2-4 Datasets of events and Observed Flow Patterns

Flow patterns are generated according to the spatio-temporal database given by Figure 2-3. All the eventsets generated from this database and the observed flow patterns are listed in Figure 2-4. If the exact locations of the events have been considered, flow patterns would not be counted as frequent patterns. If the events in the flow patterns are mapped to their relative locations according to the location of the first event in the flow, it is observed that although the exact locations of the events are not the same, they occur at the same relative locations. Event f occurs at the upper grid cell of event a and event d occurs at the right grid cell of event f in the next time interval. It can be calculated in the observed flow patterns list also. For the second flow pattern, if the location of event a is taken as the reference location which is a(1, 2), relative location of event f will be f(1-1, 3-2) which is f(0, 1). Then the locations of the events in the first and second flow patterns will be the same. It is valid for the other listed patterns. A general spatio-temporal trend is found.

Datasets of the events are stored in a sequence database. All the frequent events are found by scanning the database once. Then the set of frequent patterns are divided into partitions and projected databases are retrieved for each event in the set. And for each sequence in the projected databases, a reference location is selected. All the events are mapped to their relative locations and projected database is transformed to generalized projected database. And recursively all frequent eventsets are found by pattern growth approach. GenSTMiner algortihm is Projection-based sequential pattern mining method and smilar to the PrefixSpan[20] method.

CHAPTER 3

PROPOSED APPROACH

In this section, the proposed approach is presented in detail. The proposed approach is composed of the following four phases:

- Preprocessing
- Preparation to Mining
- Mining and Finding Frequent Rules
- Mining Rules (Second Level Mining)

Preprocessing phase, which is the first phase of the approach, is used for the preparation of the available data set for the implementation. Meteorology data is loaded into database, related database tables are created and Turkey is placed on disjoint grid cells for the ease of defining spatial neighboring relationships. Events are classified and event tables are created. Next phase is the preparation of event sets from these event tables. In this phase, another database table is created as the composition of event tables including stations' spatial information with x and y grid numbers. Next step is the construction of consecutive event sets prior to the temporal and spatial relationships in Mining and Finding Frequent Rules phase. Frequencies of the event sets are calculated and the frequent ones according to the given support are written to the output files. Mining Rules is the last phase of the approach. After all the frequent patterns are generated, for capturing the general trend of the patterns, second level mining is applied on the generated rules. In this phase, frequent consecutive event sets are classified due to their event types, intensity changes of the events according to the spatial changes are controlled and general trends are defined due to the most frequent patterns. In addition, experiments applied on the generated patterns might be considered as the analysis phase of the approach. In this phase, time is divided into three temporal sets each composed of 7 years. Frequent event changes are investigated by comparing the generated patterns of the temporal sets with each other. In the rest of this chapter, each phase is described in more detail.

3.1 Notations and Terminology

Spatio-temporal databases capture events in both time and space dimensions. Following definitions calrify the notations and terminology of the proposed approach.

Definition 1 xgridnumber, ygridnumber: The space dimension can be partitioned into a set of disjoint grid cells similar to [9, 10]. Each cell represents a spatial region and denoted as (x, y) where x is xgridnumber and y is ygridnumber.

Definition 2 Neighborhood relation: Spatial neighborhood relation is defined over the set of partitioned cells. If R represents the neighborhood relationship, Two cells (x_1, y_1) and (x_2, y_2) are said to be neighbors, denoted as $\{(x_1, y_1), (x_2, y_2)\} \in R$, if $|x_1 - x_2| \leq n_r$ and $|y_1 - y_2| \leq n_r$ where n_r is the number of grid cells.

Definition 3 Time Window: The time dimension can be divided into disjoint time windows or time periods of width W. Time t_1 is said to be temporally related to t_2 if t_1 and t_2 are in the same time window, denoted as $(t_1, t_2) \in W$. Total time is denoted as T and it is the composition of time windows. Each time window is divided into equal partitions. $W = \{w_1, w_2, ..w_i\}$ where i is the number of partitions in a time window. Total time $T = \{ w_1, w_2, ..w_i, w_{i+1},..\}$.

In this thesis T represents a year. A year is divided into 4 disjoint time windows. Each time window is composed of 3 months. Then the representation of the temporal parameters are as displayed in Figure 3-5.

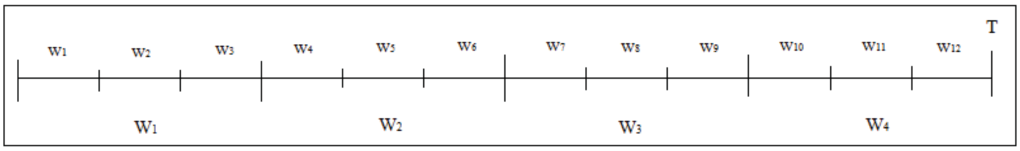


Figure 3-5 Time window representation

Definition 4 Sliding Time Window: According to the definitions, w_1 and w_2 are temporally related since $(w_1, w_2) \in W_1$. But there is not a temporal relation between w_3 and w_4 since they are not the members of the same time window. In order to relate all the consecutive time partitions, sliding window mechanism is implemented in this thesis. According to this mechanism there is a time window W that slides on T by sliding one time partition each time. wa, wb are temporalliy related if $(w_a, w_b) \in W$. Sliding window is represented in Figure 3-6.

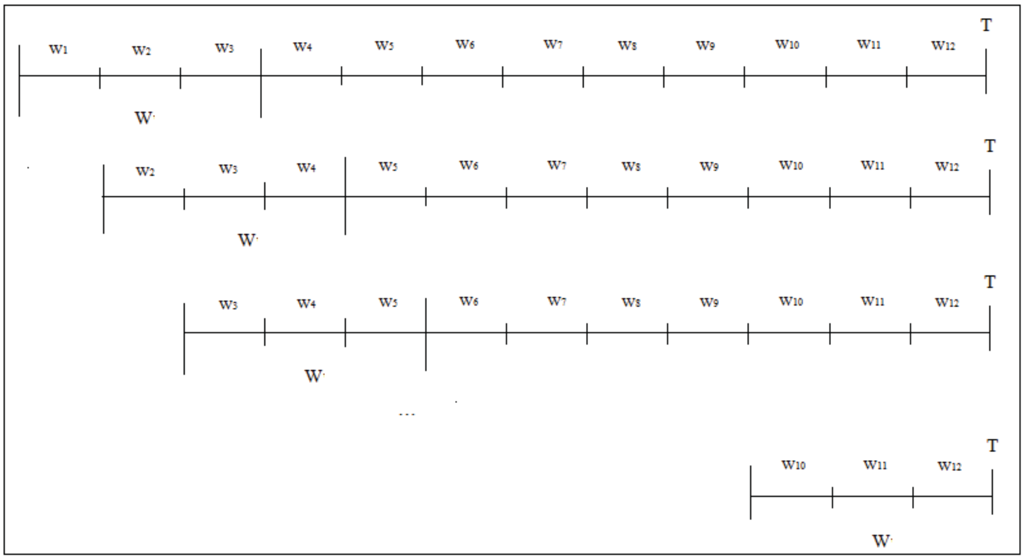


Figure 3-6 Sliding time window W

Definition 5 Eventset: An event occurring in the location (x, y) is denoted as e. An event-tuple is composed of spatial and temporal parameters and displayed as (e, T, w_i, (xgridno, ygridno)). Two events (e_1, T_1, w_1, ($xgridno_1$, $ygridno_1$)) and (e_2, T_2, w_2, ($xgridno_2$, $ygridno_2$)) are said to be related iff < ($xgridno_1$, y_1), ($xgridno_2$, $ygridno_2$)> $\in$ R and (w_1, w_2) $\in$ W, T_1 = T_2, and denoted as < ($xgridno_1$, y_1), ($xgridno_2$, $ygridno_2$)> $\in$ (R, W, T).

In this thesis, meteorological events are defined with characters. An example event-tupleset is (a, 1975, 2, (5, 7)) where a stands for the event, 1975 is T, 2 (February) is w_i, 5 is xgridno and 7 is ygridno.

Definition 6 Consecutive Eventset: A consecutive eventset is defined as a set of related events. Consecutive eventsets contain the eventsets of two consecutive time partitions of a time window. Then, the temporal relation for consecutive eventsets is (w_i, w_{i+1}) $\in$ W. Since all the eventsets are spatially and temporally related, eventsets at w_i in a consecutive eventset flow to the eventsets at w_{i+1}. Flowing property is denoted as "$\rightarrow$"

In this thesis, a time window is divided into three time partitions. Since then related eventsets are grouped into two consecutive eventsets. First consecutive eventset contains the eventsets of the first and the second time partitions, and the second consecutive eventset contains the eventsets of the second and the third time partitions. When W = {1,2,3}, eventsets (a, 1975, 1, (5, 7)) and (b, 1975, 2, (5, 6)) $\in$ (R, W, T). "1" and "2" are two consecutive time partitions of the time window W. These eventsets are members of a consecutive eventset, and eventset (a, 1975, 1, (5, 7)) flows to the eventset (b, 1975, 2, (5, 6)), denoted as (a, 1975, 1, (5, 7)) $\rightarrow$(b, 1975, 2, (5, 6)). Figured examples in the further sections illustrate spatially and temporally related eventsets and flowing patterns.

3.2 Algorithm of the Proposed Approach

Algorithm of the proposed approach is partitioned into the phases.

3.2.1 Preprocessing Phase

Input: Tables (Snow Ts, Precipitation Tp, Temperature Tt)

Output: EvenSetGrid Database table

1 Select events from Tables (Ts, Tp, Tt)

2 Classify events due to their intensities

3 Create Event Tables with the classified events (SnowEvent, PrecipitationEvent, TemperatureEvent)

4 Calculate xgidnumber and ygridnumber of all the stations

5 Create EventSetGrid table

6 Eventsets are created from Event Tables including grid numbers and inserted into EventSetGrid table

3.2.2 Preparation to Mining Phase

Input: Dt, Database Table 'EvetSetGrid'

Ts, Sample years

Ss, Sample stations

S, all stations

R, Neighborhood relationship

Output: 3 eventset vectors for sending Mining Phase

```
for each station Ssel in Ss
    for(int k=0: k=10; k++) (time window W slides 10 times)
        for each year in Ts
            vector month(k+1) = select eventsets(month(k+1), year, Ssel) from Dt
            for each station s in S,
                if( (s, Ssel) ∈R)
                vector month(k+1).add(select eventsets(month(k+1), year, s) from Dt)
                vector month(k+2) = select eventsets(month(k+2), year, Ssel) from Dt
            for each station s in S,
                 if( (s, Ssel) ∈R)
```

```
            Neighbor stations Sn, Sn.add(s)
           vector month(k+2).add(select eventsets(month(k+2), year, s) from Dt)
           vector month(k+3) = select eventsets(month(k+3), year, Ssel) from Dt
        for each station s in S,
           if( (s, Ssel) ∈R || (s, Sn) ∈R)
           vector month(k+3).add(select eventsets(month(k+2), year, s) from Dt)
        Mining(month(k+1), month(k+2), month(k+3), char indicator)
```

Note: When the indicator char is "f" it is indicated that eventsets of the last year of the Ts is sent.

3.2.3 Mining Phase

Input: vectors month(k+1), month(k+2), month(k+3)

Output: Output Files that include frequent consecutive eventsets, vectorFrequentPatterns

```
initiate vector ConsecutiveEventsets (for holding the eventsets of the 1st and the 2nd
months)
initiate vector ConsecutiveEventsets1 (for holding the eventsets of the 2nd and the 3rd
months)
while (indicator != f) (for the eventsets of each year)
    with the powerset property find all the subset eventsets of month(k+1),
month(k+2) vectors.
    Add all the eventsets ConsecutiveEventsets
    vectorsize is the size of the Consecutiveeventsets
    with the powerset property find all the subset eventsets of month(k+2), month(k+3)
vectors.
    Add all the eventsets ConsecutiveEventsets1
resultSet = FindRules(ConsecutiveEventsets , vectorSize);
resultSet1 = FindRules(ConsecutiveEventsets1 , vectorSize1);
Add resultSet and resultSet1 to vectorBlackSea
Write output files according to resultset and resultset1
OverMining(vectorFrequentPatterns)
```

Note 1: Vectorsize holds the size of the eventsets of each year. Used for the performance issues

Note 2: Consecutive Eventset has a 2 dimensional structure. First dimension holds the eventsets of a month, next dimension holds the eventsets of the consecutive month

Note 3: each ConsecutiveEventset in ConsecutiveEventsets is in the following structure;

ConsecutiveEventset.Consec1 hold the eventsets of month(k+1)

ConsecutiveEventset.Consec2 hold the eventsets of month(k+2)

each ConsecutiveEventset in ConsecutiveEventsets1 is in the following structure;

ConsecutiveEventset.Consec1 hold the eventsets of month(k+2)

ConsecutiveEventset.Consec2 hold the eventsets of month(k+3)

Note 4: resultSet and resultSet1 vectors include the frequent generated patterns

3.2.4 Frequency Calculation (FindRules())

1 for each consecutive eventset in ConsecutiveEventsets

2 search for the consecutive eventset in the eventsets of the other years

3 if consecutive eventset is frequent add eventset to resultset with its calculated percentage

4 prune resultset from the subsets of the frequent eventsets

5 return prunned resultset

3.2.5 MiningRules Phase (OverMining)

1 Initiate vectors for holding the frequent consecutive eventsets for each event type

2 add the eventsets in vectorFrequentPatterns to the classified vectors due to their event types

3 Define situation variables

4 for each consecutive eventset in the classified vector

5 control the situation and increase the count of related situation

6 write the EventChanges output file with the counts of situation variables.

3.3 Preprocessing

In this section, the first phase of the implementation is presented. Meteorology data set, which is used in the implementation, is explained, database tables created from this data set are displayed with snapshots. Classification of the events and creation of the event tables are explained in detail. Figure 3-7 is the illustration of preprocessing phase.

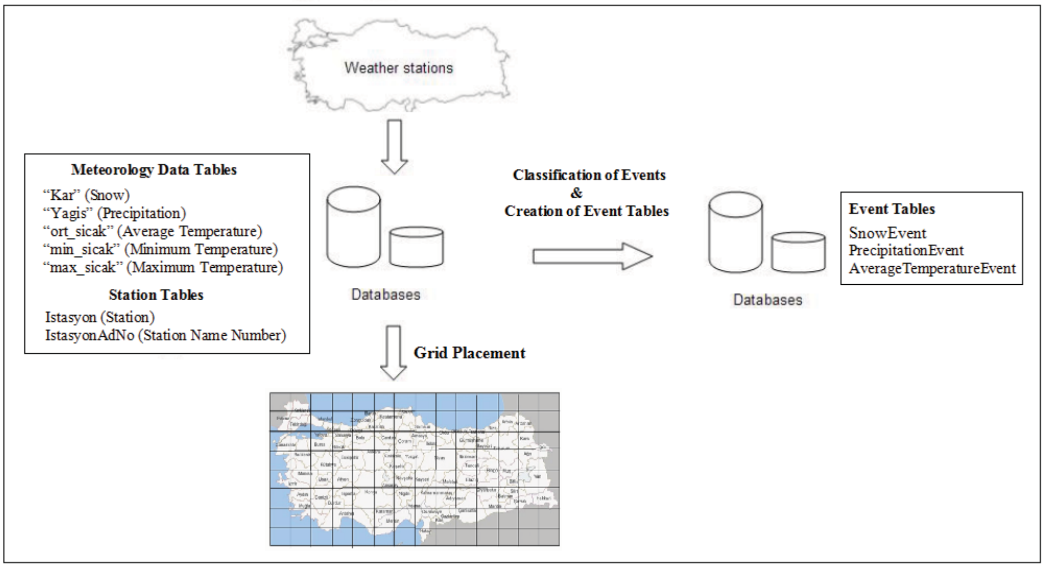


Figure 3-7 Preprocessing Phase

3.3.1 Meteorology Data Set

In this thesis, the data collected by the Turkish State Meteorological Service [16] have been used. This organization is the only legal organization providing meteorological information in Turkey. The data set covers the measurements taken from 263 major climate stations in Turkey. The monthly averages for temperature, precipitation and the number of snowy days per station from 1970 to 2007 are included in the data set. The monthly minimum and maximum values for temperature are also recorded in the stations.

The data provided by the Turkish State Meteorological Service are in text format. They are processed and written into MS-Excel documents. Then those spreadsheets are imported and gathered into database tables.

Five event tables are created named as "Kar" (Snow), "Yagis" (Precipitation), "ort_sicak" (Average Temperature), "min_sicak" (Minimum Temperature) and "max_sicak" (Maximum Temperature). Although minimum and maximum temperature data tables are created, they are not used in the experiments.

Table "Kar" consists of 15 columns which are ID, ISTASYON_ADI (Station name), YIL (Year) and 12 months from OCAK (January) to ARALIK (December) as shown in Table 3-1. Number of snowy days per month for each station is written in different rows for every different year.

Table 3-1 An Example from table “Kar” (Snow)

ID	ISTASYON_ADI	YIL	OCAK	SUBAT	MART	NISAN	MAYIS	HAZIRAN	TEMMUZ	AGUSTOS	EYLUL	EKIM	KASIM	ARALIK
1	EDIRNE	1975	1	1									7	5
2	EDIRNE	1976		3	1								1	9
3	EDIRNE	1977	8	1	1									3
4	EDIRNE	1978	2	6	1									9
5	EDIRNE	1979	20	1										
6	EDIRNE	1980	3	9	3									3
7	EDIRNE	1981	17	10										
8	EDIRNE	1982	5	2										
9	EDIRNE	1983	3	1										2
10	EDIRNE	1984		3	1									
11	EDIRNE	1985	11	16	7									
12	EDIRNE	1986	3	9	5									7
13	EDIRNE	1987	24	4	16									
14	EDIRNE	1988		1									5	4
15	EDIRNE	1989											4	8
16	EDIRNE	1990	14											
17	EDIRNE	1991	2	3										4
18	EDIRNE	1992	1	5										
19	EDIRNE	1993	11	8	6									
20	EDIRNE	1994												4
21	EDIRNE	1995	21		2								1	2
22	EDIRNE	1996	4	17	6									5
23	EDIRNE	1997			2									8
24	EDIRNE	1998	6	4	8								1	21
25	EDIRNE	1999	3	12										6
26	EDIRNE	2000	9		1									
27	EDIRNE	2001	5	1									3	28
28	EDIRNE	2002	22											5
29	EDIRNE	2003	8	8	3	2								6
30	EDIRNE	2004	14	1	1									5
31	EDIRNE	2005		13	5									11
32	EDIRNE	2006	1	11	2								1	

Precipitation and temperature data are stored in a similar way in “Yagis” (Precipitation) and “ort_sicak” (Average Temperature), tables. Monthly averages for temperature and precipitation for each station are written for each year. Example snapshots from these tables are shown in Table 3-2 and Table 3-3.

Table 3-2 An Example from table “ort_sicak” (Average Temperature)

ID	ISTASYON_NO	YIL	OCAK	SUBAT	MART	NISAN	MAYIS	HAZIRAN	TEMMUZ	AGUSTOS	EYLUL	EKIM	KASIM	ARALIK
1	17046	1977	-15.1	-10.1	-3.2	5.3	9.7	12.2	14.6	15.5	12.2	2.5	0.4	-9.5
2	17046	1978	-12.2	-7	-1.8	3.3	8.7	10.9	17.8	15.7	14	8.1	-3.2	-7.2
3	17046	1979	-9.7	-4.4	-1.9	5.6	10.1	12.5	14.8	17.8	14.2	6.7	1.9	-7.1
4	17046	1980	-12.5	-12.8	-3.8	4.2	9.9	14.3	18.6	15.2	11	5.4	2	-6.8
5	17046	1981	-11.8	-5.6	-0.2	2.6	7	13	16.9	15.7	13.5	7.5	-1.5	-5
6	17046	1982	-11.9	-12.1	-7	5.8	9.5	12.1	15.1	15.6	12	7.5		11.6
7	17046	1983	-14.1	-11.6	-2.5	5.3	9.6	12.2	16.4	15.4	10.7	5.3	1.4	-5.1
8	17046	1984	-8.8	-12.7	0.4	4.8	8.1	13.1	18.2	14.5	14.4	5.4	-0.9	11.2
9	17046	1985	-11.1	-8.6	-8.9	5.6	10.9	13.8	14.1	17.7	12	5.4	2.7	-9.2
10	17046	1986	-11.8	-8.9	-4.9	6.7	7.3	11.9	16.6	17.7	14.4	7	-1.1	-8.8
11	17046	1987	-7.6	-8.3	-8.8	-0.5	9.9	13.1	15.6	14.5	10	4	-2.2	-6.9
12	17046	1988	-13.4	-12.1	-5.1	3.7	7.8	11.8	15.4	14.5	10.4	6.1	-4.1	-5.3
13	17046	1989	-12.7	-12.6	0.4	7.6	9.5	13.3	17.4	17.5	12	6.2	-0.6	10.3
14	17046	1990	-12.1	-12.1	-6.2	3.4	8.9	13.1	16.9	15.1	12.9	6.3	-3.2	-7.5
15	17046	1991	-10.9	-12.3	-4.5	5	6.8	12.3	15	15.7	11.3	7.4		11.4
16	17046	1992	-15.6	-14.1	-8.5	1.2	7.1	11.4	14.3	15.9	11.5	7.6	-0.9	-9.1
17	17046	1993	-11	-11.9	-6.6	3.1	8.8	12.2	16.6	16.1	12.3	6.3	-4.2	-9.1
18	17046	1994	-11.1	-12.4	-3.3	7.4	9.9	12.4	16.6	15.5	14	8.1		11.1
19	17046	1995	-9.4	-9.4	-2.9	5.1	11.6	13.8	16.1	17.1	13.1	5.4	0.2	-7.5
20	17046	1996	-8.4	-9.4	-1.8	4	11.3	11.8	17.3	17.1	11.4	6.7	0.3	-0.6
21	17046	1997	-5	-7.8	-7.3	2.7	11	13	14.7	17.2	10.4	7.8	0.6	-5.8
22	17046	1998	-10.2	-10.8	-2.4	6.2	10.9	15.3	17.4	17.1	12.8	8.4	3.1	-3.5
23	17046	1999	-10.4	-6.6	-2.1	4.7	9	13.1	16.3	17.3	11.7	6.3	-0.4	-8.7
24	17046	2000	-12.7	-12.3	-8.2	6.2	8.5	12.5	19.8	17.1	12.6	6	1.1	-5.6
25	17046	2001	-12.7	-7	3	6.4	8.1	13.7	17	17.1	13.1	5.4	-1	-5.1
26	17046	2002	-12.7	-8.3	-1.7	3.3	8.4	12.6	15.9	14.9	13.3	7.9	-1.6	10.7
27	17046	2003	-9.2	-7.5	-6.3	3.9	10.6	12.4	16	16.6	11.7	8.5	-0.7	-8.7
28	17046	2004	-10.7	-6.7	-1	3.7	8.9	12.7	15.4	16.7	12.1	7.4		12.2
29	17046	2005	-13.1	-12.5	-3	5.4	9.7	12.4	17.9	17.1	11.8	5.9	0.4	-5.6
30	17046	2006	-11.7	-6.7	-0.7	5.6	10.7	16	16.2	19.8	12.7	8.4		12
31	17085	1975	2.1	3.5	10.3	15.7	17.5	23.1	25.2	23.9	19.9	14	8.2	1.7

Table 3-3 An example from table “Yagis” (Precipitation)

ID	ISTASYON_NO	YIL	OCAK	SUBAT	MART	NISAN	MAYIS	HAZIRAN	TEMMUZ	AGUSTOS	EYLUL	EKIM	KASIM	ARALIK
1	17986	1975	135.6	159.3	100.8	91.6	12.7	4.2		0.6	35.9	101.1	180.9	165.3
2	17986	1976	211.6	63.2	100.5	140.8	225	7.1	8.1	5.1	67.2	154.1	157.6	128.3
3	17986	1977	154.7	79.6	95.8	86.8	14.6	5.3			7.6	68.9	11.5	186.6
4	17986	1978	305.2	88.2	88.1	33.4		63.2	0.7		68.3	160.1	3.7	168.6
5	17986	1979	104.3	102.4	69.4	47.5	29.5		3.4		1.3	206.2	109.8	123.7
6	17986	1980	166.5	112.2	135	88	0.8	0.3		0	4.7	125.9	68.3	121.3
7	17986	1981	279.5	137.7	115.6	51.4	10.5	30.3			12.7	17.9	91.9	167.9
8	17986	1982	131.9	58.5	134.1	62.1	7.2	18.3	19.8		3	170	36	76.8
9	17986	1983	52.5	168.9	150.4	203.9	25.6	0.2	2.2	32.4	118.2	63.6	287.6	56.9
10	17986	1984	150.2	120.7	194.5	193	0.3	1.2	14	5.8		14	68.5	64.3
11	17986	1985	175.6	211.1	18.9	40.2	1.8	0.5		0	4.4	145.6	29.9	60.4
12	17986	1986	106.5	75	78.2	23.8	87.5	9.9			5.3	134.3	169.4	96.7
13	17986	1987	223.5	131.1	204	9.5	15.1	2	24.5	1.4	1.1	103	98.2	154.3
14	17986	1988	128.9	153.6	169.8	25.7	94.3	45.6		1.6	8.9	153	161.2	129
15	17986	1989	29.1	11.2	107.8	3.9	31.7	13.4			52.5	97.6	149	90
16	17986	1990	94.6	159.2	20.7	21.6	83.9	22.8	4.8		3.8	55.1	35.2	68.3
17	17986	1991	111.4	120.6	88.9	61.1	42.5		1.4		2.3	95.6	84.4	433.6
18	17986	1992	58.2	152.2	52.5	64.2	299.2	50.5	40.2	0	92.4	16.5	232.2	188.6
19	17986	1993	32.6	73	87.9	21.5	85.5	54		0.3	2	2.7	66.9	40.4
20	17986	1994	213.3	241	45.4	39.4	10.6	10.4	82.5	10.8	0.2	53.6	151.1	141.7
21	17986	1995	147.9	85.5	74.6	61.1	59.3	27.4	18.7	0.3	61.8	26.4	189.3	31.2
22	17986	1996	97.4	72.3	236.9	116.8	3.8	9	0		44.6	220.1	14.1	192.8
23	17986	1997	58.7	121.9	77.4	120.7	22.1			3.6	52.8	289.2	168.7	169.8
24	17986	1998	78.8	102.8	282.2	40.7	51.9	0	0.2		118.6	28.6	135.3	239.2
25	17986	1999	61	104.6	76.5	81.8	0.3	2.8		5.7	89.2	94.3	23.4	92.2
26	17986	2000	304.8	163.6	52.1	62.3	2.5				116.9	64.9	67.6	133.4
27	17986	2001	68.2	185.1	73.9	40.3	43.7	0.2		24.2	51.1	121.7	115.9	274.3
28	17986	2002	234.9	35.7	100.6	78.6	55.6	22.7	2.8	115.1	53.9	42.3	90	210.3
29	17986	2003	102.3	194.8	213.5	24.8	27.1	1.5			6.7	26.9	132.9	189.7
30	17986	2004	199.2	110.6	43.8	41.2	5.1	8.5		15.8		60.1	179.8	62.9
31	17986	2005	67.3	89.9	52.1	40.1	31.3	53.8	11	0.3	213.2	84.2	105.6	142.2
32	17986	2006	126.6	110.4	112.3	19.9	0.2		8.6		138.2	235.4	115.4	22.7

“Istasyon” (Station) table contains the station number, X and Y coordinates, and Z column as the height of the station. An example snapshot from this table is displayed in Table 3-4.

Table 3-4 An Example from table "Istasyon" (Station)

ISTASYON_N0	Y	X	Z
17022	41.45	31.8	137
17030	41.283333	36.3	4
17033	40.983333	37.9	4
17034	40.916667	38.383333	37
17037	41	39.716667	30
17040	41.033333	40.516667	9
17042	41.4	41.433333	33
17045	41.183333	41.816667	628
17050	41.666667	26.566667	51
17052	41.683333	27.3	174
17052	41.733333	27.233333	232
17056	40.983333	27.55	3
17059	41.25	29.033333	30
17069	40.783333	30.416667	30
17070	40.733333	31.6	743
17074	41.366667	33.783333	800
17080	40.6	33.616667	751
17083	40.866667	35.466667	755
17084	40.55	34.966667	776
17085	40.65	35.85	412
17086	40.3	36.566667	608
17088	40.466667	39.466667	1219
17090	39.75	37.016667	1285
17096	39.916667	41.266667	1758
17099	39.733333	43.05	1632
17100	39.916667	44.05	858
17110	40.183333	25.9	72
17111	39.833333	26.066667	28
17112	40.133333	26.4	6
17114	40.35	27.966667	58
17116	40.183333	29.066667	100

"IstasyonAdNo" (Station Name Number) is the table that is created for holding the station numbers and corresponding station names. Such a table is necessary since some of the tables include station names and some of them include station numbers. An example from this table is shown as Table 3-5.

Table 3-5 Example instances from table ”IstasyonAdNo”

ID	Istasyon_no	Istasyon_adi
1	17045	ARTVIN
2	17046	ARDAHAN
3	17085	AMASYA
4	17099	AGRI
5	17130	ANKARA
6	17175	AYVALIK
7	17184	AKHISAR
8	17190	AFYON
9	17192	AKSARAY
10	17234	AYDIN
11	17239	AKSEHIR
12	17265	ADIYAMAN
13	17300	ANTALYA
14	17310	ALANYA
15	17320	ANAMUR
16	17351	ADANA
17	17372	ANTAKYA
18	17602	AMASRA
19	17612	AKCAKOCA
20	17626	AKCAABAT

There are a total of 4972 records in the raw data of table “Kar”, 8281 records in table “Yagis” and 8242 records in table “ort_sicak”.

3.3.2 Creation of Event Tables

In this phase, events are classified on the basis of their values, and related event tables are created for each of the event type. Snow events are classified as shown in Table 3-6.

Table 3-6 Snow Event Classification

Number of Snowy Days	Event
snowyDays <= 5	g
5 <= snowyDays <=9	h
9 <= snowyDays <=13	i
13 <= snowyDays <=17	j
17 <= snowyDays <=21	k
snowyDays >= 21	l

Table "SnowEvent" is created for holding the records that have specific snow events. A sample snapshot of the table "SnowEvent" is shown as Table 3-7.

Table 3-7 An example for table "SnowEvent"

istasyon_no	yil	ay	event
17050	1975	1	g
17050	1975	2	g
17050	1975	3	g
17050	1975	4	g
17050	1975	5	g
17050	1975	6	g
17050	1975	7	g
17050	1975	8	g
17050	1975	9	g
17050	1975	10	g
17050	1975	11	h
17050	1975	12	g
17050	1976	1	g

Since the records are held for each station, for each year and for each month separately, the size of the table "SnowEvent" is about 12 times bigger than table "Kar" and contains 59664 rows of crisp data.

A similar approach is followed for the temperature and precipitation events. Precipitation events are classified according to Table 3-8.

Table 3-8 Precipitation Event Classification

Precipitation	Event
precipitation <= 25.0 mm	1
25.0 mm <= precipitation <=50.0 mm	2
50.0 mm <= precipitation <=75.0 mm	3
75.0 mm <= precipitation <=100.0 mm	4
100.0 mm <= precipitation <=125.0 mm	5
precipitation >= 125.0 mm	6

Table “PrecipitationEvent” is created for holding the records that have specific precipitation events. A sample part of the table “PrecipitationEvent” is shown as Table 3-9.

Table 3-9 An example for table “PrecipitationEvent”

istasyon_no	yil	ay	event
17986	1975	1	6
17986	1975	2	6
17986	1975	3	5
17986	1975	4	4
17986	1975	5	1
17986	1975	6	1
17986	1975	7	1
17986	1975	8	1
17986	1975	9	2
17986	1975	10	5
17986	1975	11	6
17986	1975	12	6
17986	1976	1	6

Table “PrecipitationEvent” contains 99225 rows of crisp data. Average Temperature events are classified due to the values given in Table 3-10.

Table 3-10 Average Temperature Event Classification

Average Temperature	Event
AvgTemp <= 5.0 C°	a
5.0 C°<= AvgTemp <=10.0 C°	b
10.0 C°<= AvgTemp <=15.0 C°	c
15.0 C°<= AvgTemp <=20.0 C°	d
20.0 C°<= AvgTemp <=25.0 C°	e
AvgTemp >= 25.0 C°	f

Table “AverageTemperatureEvent” (Average Temperature) is created in db1 for holding the records that have specific temperature events. A sample part of the table “AverageTemperatureEvent” is shown as Table 3-11.

Table 3-11 An example for table "AverageTemperatureEvent"

istasyon_no	yil	ay	event
17046	1977	1	a
17046	1977	2	a
17046	1977	3	a
17046	1977	4	b
17046	1977	5	b
17046	1977	6	c
17046	1977	7	c
17046	1977	8	d
17046	1977	9	c
17046	1977	10	a
17046	1977	11	a
17046	1977	12	a
17046	1978	1	a

Table "AverageTemperatureEvent" contains 98945 rows of crisp data.

Classes and methods that are implemented for the creation of event tables are explained in implementation details section with snapshots of the sample codes.

3.3.3 Grid Placement

The space dimension is partitioned into a set of disjoint grid cells as in [10]. Each cell represents a spatial region (or location), denoted as (x, y). Weather stations are defined with their general properties as X coordinate, Y coordinate, station number, station name and height of the station. In our implementation, X and Y coordinates of the stations are converted into *x grid numbers* and *y grid numbers*. Neighborhood relationships which are used for spatial associations are defined easily with the help of this representation.

Any value can be given for defining the neighborhood relationship. In this study, it is assumed that if the x grid number and y grid number difference of two stations is smaller than or equal to 1, these stations are neighbors. As an example, Station1 (2,3) and Station2 (3,4) are assumed to be neighbors since the difference of their x and y grid numbers is 1. Calculation of neighboring relationships of stations with respect to their grid numbers by the implementation is explained in the implementation details section.

3.4 Preparation to Mining

This is the second phase of the approach. Preparation of the eventsets for mining frequent patterns is the main purpose of this phase. Creation of eventsets from the related table is explained in detail in the following subsection. Afterwards, an eventset preparation example is given for a better understanding of the phase.

3.4.1 Preparing Eventsets

As the first task of this phase, named “EventSetGrid” table is created. According to the station number or station name of each station, events are selected from tables SnowEvent, PrecipitationEvent, AverageTemperatureEvent. All the selected records are inserted into the table “EventSetGrid”. This table contains the information of each station with its xgridno and ygridno and for all the months of each year with the related events of each type. As a summary, table “EventSetGrid” is the result of preprocessing phase. An example snapshot from this table is given in Table 3-12.

Table 3-12 An example from table “EventSetGrid”

istasyon_no	xgridno	ygridno	yil	ay	event
17022	11	1	2003	12	g
17022	11	1	2005	3	g
17022	11	1	2005	4	g
17022	11	1	2005	5	g
17022	11	1	2006	1	h
17022	11	1	2006	2	j
17022	11	1	2006	3	g
17022	11	1	2006	4	g
17022	11	1	2002	4	g
17022	11	1	2002	5	g
17022	11	1	2002	6	g
17022	11	1	2003	2	k
17022	11	1	2003	3	i
17022	11	1	2003	4	g
17022	11	1	2003	5	g
17022	11	1	2004	8	g
17022	11	1	2004	9	g
17022	11	1	2004	10	g
17022	11	1	2005	6	g
17022	11	1	2005	7	g
17022	11	1	2005	8	g
17022	11	1	2005	9	g
17022	11	1	2006	12	h
17022	11	1	2000	11	1
17022	11	1	2000	12	6
17022	11	1	2001	1	3
17022	11	1	2001	2	5
17022	11	1	2001	3	5
17022	11	1	2001	4	4
17022	11	1	2001	12	6
17022	11	1	1975	3	4

Table “EventSetGrid” contains 204637 rows of crisp data which is the total of the records in SnowEvent, PrecipitationEvent and OrtSicaklikEvent tables.

There are too many stations, and execution time takes very long if the developed application is run for all of the stations. Therefore,   a sample set of stations is chosen for observing the outputs of the program. Sample set contains the stations in Black Sea Region.

Due to its distinct geographical conditions (i.e. variation of altitude, orientation of mountains, coastal effects, etc.) climatic changes can be observed significantly within Black Sea Region, especially between Eastern and Western parts or coastal and inland parts of the region. List of the sample stations with their station numbers is given in Table 3-13.

Table 3-13 Set of the Sample stations

Station Number	Station Name
17045	Artvin
17602	Amasra
17626	Akcaabat
17034	Giresun
17088	Gumushane
17042	Hopa
17033	Ordu
17030	Samsun
17040	Rize
17037	Trabzon
17022	Zonguldak
17624	Unye
17622	Bafra
17646	Cerkes
17656	Arpacay

Similar to the approach given in [9], time windows are defined. A year is divided into 4 time windows and each window is composed of 3 months. Events (they are represented in eventsets structure in the implementation which holds the information of station number, year and the month of the event occurred, a character to specify the event, x grid number and y grid number of the station) of a specified station and events of neighbor stations (neighbor of the specified station) for the first month of the time window are found. The same procedure is followed for the second month of the time window. Finally, for the third month of the time window, events of all the neighbor stations of the selected station, besides events of all the neighbor stations of the neighbor stations (neighbors of the neighbors of selected station) are found. These three months' sets are prepared for the mining phase. Associations within the same time windows are investigated.

As mentioned in the previous sections a sliding mechanism is proposed in the approach. Sliding through the time windows is possible. Outputs are given for the months in the order of January-February-March, February-March-April, March-April-May,,,,October-November-December. By this way, the loss of any relations between different time windows is prevented.

3.4.2 Eventset Preparation Example

It will be beneficiary to illustrate the implementation with Figure 3-8.

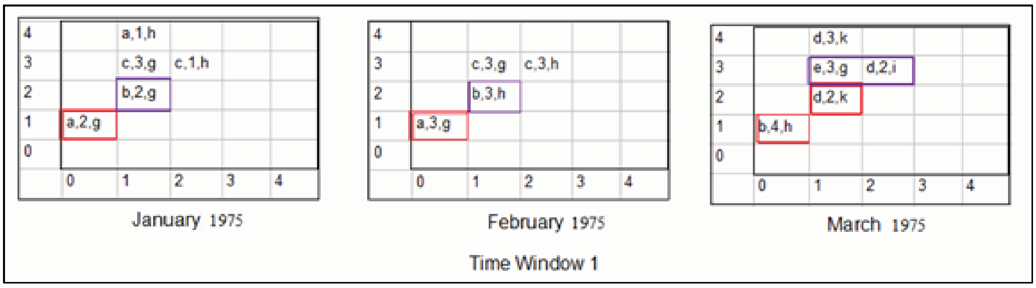


Figure 3-8 Eventset Preparation example

For the example illustrated Figure 3-8, it is assumed that there are weather stations, in the numbered grid cells (0, 1), (1, 2), (1, 3), (1, 4) and (2, 3). Characters and numbers written in the grid cells stand for the events. Since there are three types of events, each of them represents a type of event. Events change by the change of the months. As told in the previous sections, each time window is composed of 3 months. Eventsets created from the example can be listed as follows:

January	February	March
a, 1975, 1, (0,1)	a, 1975, 2, (0,1)	b, 1975, 3, (0,1)
2, 1975, 1, (0,1)	3, 1975, 2, (0,1)	4, 1975, 3, (0,1)
g, 1975, 1, (0,1)	g, 1975, 2, (0,1)	h, 1975, 3, (0,1)
b, 1975, 1, (1,2)	b, 1975, 2, (1,2)	d, 1975, 3, (1,2)
2, 1975, 1, (1,2)	3, 1975, 2, (1,2)	2, 1975, 3, (1,2)
g, 1975, 1, (1,2)	h, 1975, 2, (1,2)	k, 1975, 3, (1,2)
c, 1975, 1, (1,3)	c, 1975, 2, (1,3)	e, 1975, 3, (1,3)
3, 1975, 1, (1,3)	3, 1975, 2, (1,3)	3, 1975, 3, (1,3)
g, 1975, 1, (1,3)	g, 1975, 2, (1,3)	g, 1975, 3, (1,3)
c, 1975, 1, (2,3)	c, 1975, 2, (2,3)	d, 1975, 3, (2,3)
1, 1975, 1, (2,3)	3, 1975, 2, (2,3)	2, 1975, 3, (2,3)
h, 1975, 1, (2,3)	h, 1975, 2, (2,3)	i, 1975, 3, (2,3)
a, 1975, 1, (1,4)		d, 1975, 3, (1,4)
1, 1975, 1, (1,4)		3, 1975, 3, (1,4)
h, 1975, 1, (1,4)		k, 1975, 3, (1,4)

Next step is the creation of month1, month2 and month3 vectors (structure defined in the implementation for holding the events). The selected station is the one that is located in a red box in Figure 3-8. All the eventsets of this station and eventsets of its neighbor stations are added to month1 vector. In the figure, selected station has only one neighbor with events in it (shown with the color of magenta). Since then vector1 is composed of the eventsets of this selected station's and its neighbor station's listed in the January column which are given below:

a, 1975, 1, (0,1)
2, 1975, 1, (0,1)
g, 1975, 1, (0,1)
b, 1975, 1, (1,2)
2, 1975, 1, (1,2)
g, 1975, 1, (1,2)

Elements of month2 are found in a similar way. Eventsets of the selected station and its neighbor stations in the following month (in the example it is February) are found and added to month2 vector. Month2 vector is composed of the following eventsets;

a, 1975, 2, (0,1)
3, 1975, 2, (0,1)
g, 1975, 2, (0,1)
b, 1975, 2, (1,2)
3, 1975, 2, (1,2)
h, 1975, 2, (1,2)

Month3 vector includes the eventsets of the selected station and its neighbor stations. Additionally it contains the information from the neighbor stations of its neighbor stations in the following month (in the example it is March). Selected station is in the numbered grid cell (0,1). It has one neighbor station that includes events in it, and it is in the numbered grid cell (1,2). This time, neighbors of neighbor stations are also considered. And according to the above example, these stations are the ones in the numbered grid cells (1,3) and (2,3).

Therefore, month3 vector is composed of the following eventsets;

b, 1975, 3, (0,1)
4, 1975, 3, (0,1)
h, 1975, 3, (0,1)
d, 1975, 3, (1,2)
2, 1975, 3, (1,2)

k, 1975, 3, (1,2)
e, 1975, 3, (1,3)
3, 1975, 3, (1,3)
g, 1975, 3, (1,3)
d, 1975, 3, (2,3)
2, 1975, 3, (2,3)
i, 1975, 3, (2,3)

Another point that should be highlighted is the selection of the years for building eventsets. Above example is illustrated only for the year 1975. However in the real application, during the building of month vectors, the data for all years are taken into account. The monthly events for the corresponding years are investigated. Due to performance issues, only the data between 1975 and 1991 are used in the experiments.

3.5 Mining and Finding Frequent Rules

This chapter describes the mining technique used for finding frequent spatio-temporal rules. Additionally, techniques that are used to improve the performance of the implementation are explained.

3.5.1 Mining Phase

A vector structure is defined for holding the consecutive events. With a two dimensional structure, it is possible to hold the events of two consecutive months. Eventsets within a time-window are added into two vectors. One of the vectors holds the events of the first and the second months of the time-window, the other vector holds the events of the second and the third months of the time window. All of the calculated subsets are stored in these vectors and the vectors that include the eventsets are ready to be mined.

After all the subsets of the eventsets for the mentioned 16 years are kept in the defined vector structures, next step is finding the frequencies of these eventsets.

Again, it is beneficiary to illustrate the mining phase with an example. To this aim, eventsets of the previous example (eventset Preparation example) will be used. All the eventsets of the first time window are as listed in Figure 3-9.

eventsets of month1	eventsets of month2	eventsets of month3
a, 1975, 1, (0,1)	a, 1975, 2, (0,1)	b, 1975, 3, (0,1)
2, 1975, 1, (0,1)	3, 1975, 2, (0,1)	4, 1975, 3, (0,1)
g, 1975, 1, (0,1)	g, 1975, 2, (0,1)	h, 1975, 3, (0,1)
b, 1975, 1, (1,2)	b, 1975, 2, (1,2)	d, 1975, 3, (1,2)
2, 1975, 1, (1,2)	3, 1975, 2, (1,2)	2, 1975, 3, (1,2)
g, 1975, 1, (1,2)	h, 1975, 2, (1,2)	k, 1975, 3, (1,2)
		e, 1975, 3, (1,3)
		3, 1975, 3, (1,3)
		g, 1975, 3, (1,3)
		d, 1975, 3, (2,3)
		2, 1975, 3, (2,3)
		i, 1975, 3, (2,3)

Figure 3-9 Eventsets of the time window

As mentioned previously, eventsets of a time window are added to two vectors. First of these two vectors includes the eventsets of month1 and month2, second one includes the eventsets of month2 and month3. Each vector is defined in a two dimensional structure to hold the eventsets of a month and eventsets of the following month. All the eventsets hold by these two vectors (consecutive eventsets) are listed in Table 3-14.

Table 3-14 Consecutive eventsets of the time window

First Consecutive eventset of the time window	{(a, 1975, 1, (0,1)), (2, 1975, 1, (0,1)), (g, 1975, 1, (0,1)), (b, 1975, 1, (1,2)), (2, 1975, 1, (1,2)), (g, 1975, 1, (1,2)) }
	{(a, 1975, 2, (0,1)), (3, 1975, 2, (0,1)), (g, 1975, 2, (0,1)), (b, 1975, 2, (1,2)), (3, 1975, 2, (1,2)), (h, 1975, 2, (1,2))}
Second Consecutive eventset of the time window	{(a, 1975, 2, (0,1)), (3, 1975, 2, (0,1)), (g, 1975, 2, (0,1)), (b, 1975, 2, (1,2)), (3, 1975, 2, (1,2)), (h, 1975, 2, (1,2))}
	{(b, 1975, 3, (0,1)), (4, 1975, 3, (0,1)), (h, 1975, 3, (0,1)), (d, 1975, 3, (1,2)), (2, 1975, 3, (1,2)), (k, 1975, 3, (1,2)), (e, 1975, 3, (1,3)), (3, 1975, 3, (1,3)), (g, 1975, 3, (1,3)), (d, 1975, 3, (2,3)), (2, 1975, 3, (2,3)), (i, 1975, 3, (2,3))}

If the eventsets are projected with the event characters for the ease of understanding, same table will be structured as Table 3-15.

Table 3-15 Events of the consecutive eventsets

First Consecutive eventset of the time window	{ a, 2, g, b, 2, g }
	{ a, 3, g, b, 3, h }
Second Consecutive eventset of the time window	{ a, 3, g, b, 3, h }
	{ b, 4, h, d, 2, k, e, 3, g, d, 2, i }

Flow of events in the first consecutive eventset is displayed as; { a, 2, g, b, 2, g } → { a, 3, g, b, 3, h }. First set in the consecutive eventset includes the events of a station and events of its neighbor stations for January in 1975, and the second set in the consecutive eventset includes the events of a station and events of the neighbor stations in February in 1975. Next step is calculating the frequencies of the events in these eventsets by searching the consecutive eventset in the other years. Frequency count is set to 10. This corresponds to about 70% support threshold. Since there are 16 years to look for frequent eventsets, if an eventset occurs in at least 10 years out of 16, it is assumed that this eventset is frequent. All the events in the consecutive eventset might not be frequent. For finding the frequent ones, subsets of the consecutive eventsets are calculated. For the first consecutive eventset, subsets are; {{a} → {a}, {a} → {3}, {a} → {g}, {a} →{b}, {a} →{3}, {a} →{h}, {a, 2} →{a}, {a, 2} →{3}, {a, 2} →{g}…}.

While trying to find out if the eventsets are frequent or not, many properties of the eventsets are compared with each other. There are also some performance issues for getting faster results in the implementation. Structure of the vectors, how the consecutive events are hold, how their frequencies are calculated are explained in the implementation details section.

Once the frequencies of the eventsets are calculated, the frequent ones are hold in result sets. However as mentioned previously, subsets of the eventsets are found and sent to the next phase for finding the frequent sets. If a set is frequent all of its subsets are frequent either. Due to this property, found result sets which hold the frequent eventsets include both the frequent sets and all their subsets. For example, in the above example, if {a, 2} → {3} is found as a frequent eventset, {a} → {3} which is a subset of the previous eventset is found as a frequent eventset either. {a} → {3} eventset should be eliminated from the frequent eventsets. Another section "Pruning" phase is implemented to eliminate the subsets of the frequent sets from the result set.

After all the frequent eventsets are found, output files can be prepared. Output Files are written separately for each of the stations. Unique file names are defined for each of the output files. The station number, numeric representation of the month or months and some dashes for splitting the characters form the file name. For example; "17045--month--1" is the file name that is prepared for the result set of station 17045 (Artvin), includes the frequent eventsets for the first and the second (January and February) months. File name "17045--1--ay—2" indicates that output file is prepared for the station numbered 17045 and file includes eventsets for the second and the third months. Content of the output files, how they are interpreted will be explained in the further sections.

While writing the frequent eventsets to the output files, each frequent eventset is added to another structure which is in the type of a vector. At the end of the writing process, this vector (vector Karadeniz) includes all the frequent eventsets for all of the stations and for all of the years. Then these frequent eventsets are processed in another phase named "Mining Rules".

3.6 Mining Rules

Mining Rules is the last phase of the implementation. This phase is developed in order to find new and higher-level rules from the generated patterns. It is the generalization of the patterns and can be called as *"second level mining"*.

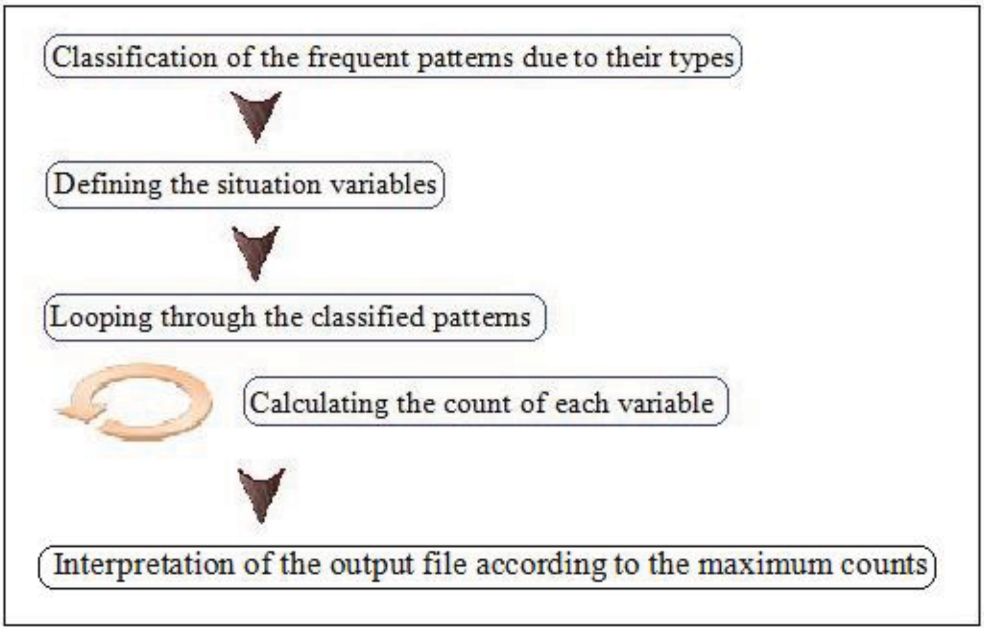


Figure 3-10 Second Level Mining

Figure 3-10 illustrates the modality of Mining Rules phase. Since the defined approach is based on the eventsets which are composed of events, first step of this phase is the classification of the frequent patterns on the basis of their types (event types). An important point that should be considered is the usage of frequent patterns in this phase. It is the reason why the rules are the generalization of the generated patterns. Next step is the definition of situation variables. These variables might be chosen to define the temporal or spatial relationships. The purpose is to find out the general trend of the frequent patterns according to the spatial and/or temporal relationships such as the reflection of the temporal and spatial changes on the events. After the situations are defined, all the frequent patterns which are classified on the basis of their types are overviewed and events in the evensets are compared with each other according to the defined parameters. Counts of the situation variables are calculated as a result of these comparisons. Next is the interpretation of the output file. It is assumed that the most frequent situation (the situation variable with the maximum value) shows the general trend.

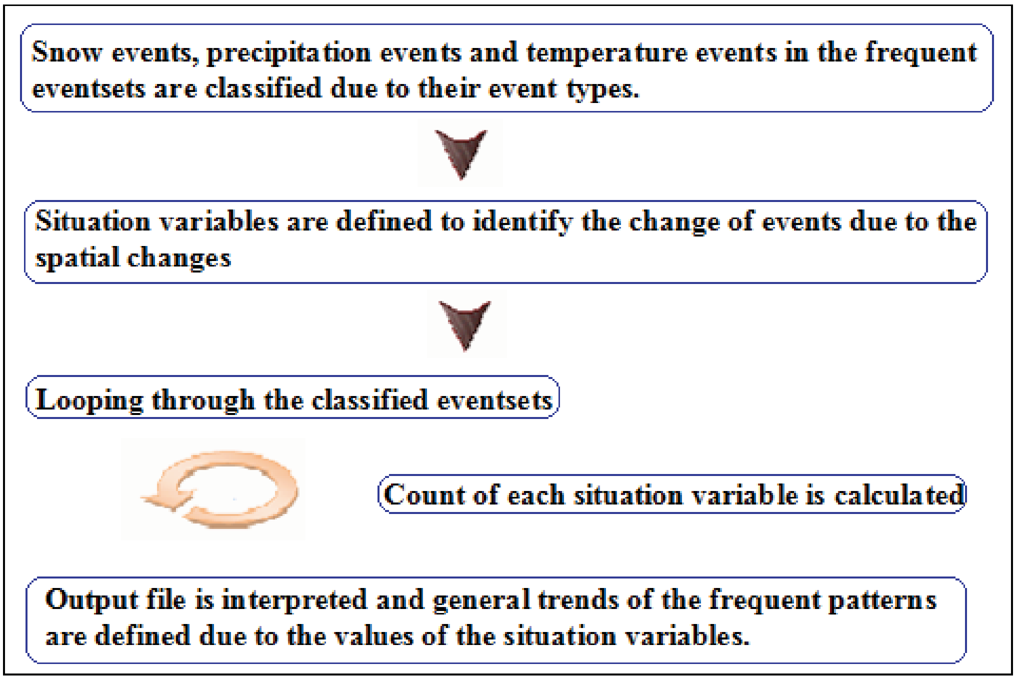


Figure 3-11 Second level mining on the available data set

Figure 3-11 illustrates the second level mining phase. Events are grouped according to their event types as; snow events, precipitation events and temperature events in this phase. For each event type 4 situations are defined for representing event changes due to x grid number changes and 4 situations are defined for representing event changes due to y grid number changes. Variables are initiated to count the number of the elements for each defined situation. Characters which represent the events are converted into numeric values from 1 to 6 according to their intensities in the increasing order. For example, temperature event represented by 6 is hotter than temperature event represented by 1. Numerical illustration of the events facilitates comparison of the intensities of the events.

Within the frequent consecutive eventsets, numeric values of the events are compared. Another comparison is between the grid numbers of the events. For example, if the event number of a temperature event that occurred in a station is greater than the event number of an event that occurred in another station in the frequent eventset and the x grid number of the station is greater than the x grid number of the compared station then the variable 'sicaklikGridUpEventUpCountxgridno' is increased by one while controlling all the eventsets that contain temperature events. In other words, temperature increases by the increase of the x grid number.

After all the eventsets are controlled and variable counts are determined, it is possible to interpret the change of the events due to the change of grid numbers.

Values of the parameters and short descriptions for defining what these parameters stand for are written to the output file named "EventChanges".

An important point that should be considered in this phase is the spatial relationship of the events which are compared with each other. Stations in the frequent eventsets are associated with a neighboring relationship. From this point of view, it is not possible to generate a rule for the stations which do not have a neighboring relationship. General trends like *'stations located up North have higher average temperatures compared to the average temperature of the stations located to the South of them'* can be captured by the interpretation of the values. Implementation details of this phase are given in the 'Implementation Details' section and interpretation of the output file can be found under 'Results of Mining Rules Phase' section.

CHAPTER 4

IMPLEMENTATION DETAILS

This section presents the implementation details in the order of the previous section.

4.1 Preprocessing

Preprocessing phase is composed of the creation of event tables and grid placement. These functionalities are explained in detail from the implementation point of view in the further sections.

4.1.1 Creation of Event Tables

EventMaker package displayed in Figure 4-12 includes the java classes for the classification of the events according to their intensity and creates three event tables named as "SnowEvent", "PrecipitationEvent" and "AverageTemperatureEvent".

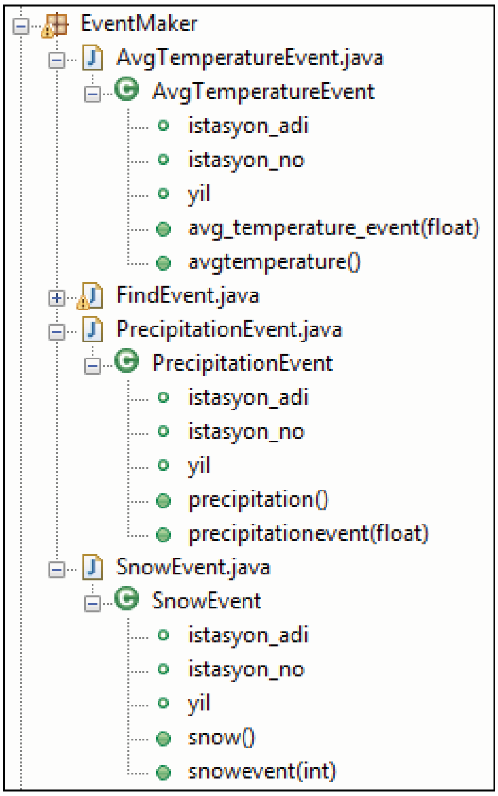


Figure 4-12 EventMaker package

In FindEvent.java class, instances of SnowEvent, AvgTemperatureEvent and PrecipitationEvent are created and related methods of these classes are called throughout the body of this class.

In *snow()* method of SnowEvent.java class all records from table “Kar” (Snow) are read, number of snowy days is sent to *snowevent()* method for the classification of the event and events are classified due to the “Snow Event Classification” table. Table “SnowEvent” is created for holding the records that have specific snow events defined by the *snowevent()* method.

In *precipitation()* method of PrecipitationEvent.java class, all records from table “Precipitation” are read, value of average precipitation is sent to *precipitationevent()* method for the classification of the event and events are classified due to the “Precipitation Event Classification” table. Table “Precipitation Event” is created in db1 for holding the records that have specific precipitation events defined by the *precipitationevent()* method.

In *avgtemperature()* method of AvgTemperatureEvent.java class, all records from table “ort_sicak” are read, value of average temperature is sent to *avg_temperature_event()* method for the classification of the event and events are classified due to the “Average Temperature Event Classification” table. Table “AverageTemperatureEvent” is created in db1 for holding the records that have specific temperature events defined by the *avg_temperature_event()* method.

4.1.2 Grid Placement

Istasyon.java class is written for clarifying the general properties of stations. X coordinate, Y coordinate, station number, station name and height of the station are defined in this class.

GridStations.java class is implemented for reading all the station records from table “Istasyon” (Station) by the *ReadIstasyon()* method and setting the properties of stations by creating objects from Istasyon.java class.

All station objects are kept in a vector and this vector is sent to the *returnGridStations()* method. For each of the station objects kept in the vector, a new object is created from Grid.java class.

Grid.java

Grid.java class has similar properties as Station.java class. But instead of x and y coordinates, Grid.java class has xgridno and ygridno properties for holding the location information in terms of gridno instead of coordinates.

Grid numbers of the stations are calculated by the methods *FindGridNoX()* and *FindGridNoY()* in GridStations.java class while creating and assigning Grid objects for each of the stations. Parameters and methods of GridStations.java class are displayed in Figure 4-13.

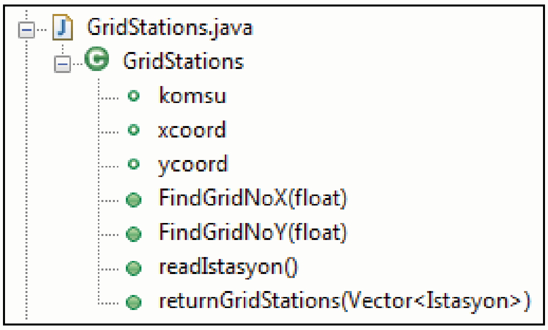


Figure 4-13 GridStations.java class

returnGridstations() method returns a vector that contains all of the created Grid objects. Defined vectors and the return statement of the method can be seen in Figure 4-14.

```
public static Vector<Istasyon> Stations = new Vector<Istasyon>();
public static Vector<Grid> gridStations = new Vector<Grid>();

    GridStations findStations = new GridStations();
    Stations = findStations.readIstasyon();
    gridStations = findStations.returnGridStations(Stations);
```

Figure 4-14 Implementation of vectors and methods

4.2 Preparation of Mining

Preparation of Mining phase which is the second phase of the implementation includes the preparing eventsets functionality. This functionality is explained in detail with source code examples and structure definitions in the further sections.

Preparing Eventsets

Next step is creating the eventsets for mining. Implementation of eventset creation and related method call is displayed in Figure 4-15.

```
GridEventSet createGridEventSet = new GridEventSet();
createGridEventSet.FindEvensets(gridStations);
```

Figure 4-15 Implementation of eventset creation

GridEventSet.java is implemented for the creation of table “EventSetGrid” and for inserting the eventsets into the table. Methods of GridEventSet.java class are shown in Figure 4-16.

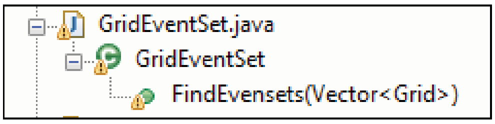


Figure 4-16 Properties and Methods of GridEventSet.java

gridStations vector returned by *returnGridStations()* method is sent to *FindEventsets()* method of this class. And according to the station number or station name of each element in the sent vector, events are selected from tables KarEvent, YagisEvent, OrtSicaklikEvent. All the selected records are inserted into the table “EventSetGrid”. This table contains the information of each station with its xgridno and ygridno and for all the months of each year with the related events of each type.

GstMiningPrep.java

gridStations vector which is returned by the *returnGridStations()* method is sent to *mineEvents()* method of GstMiningPrep.java class. Implementation of *mineEvents()* method call is displayed in Figure 4-12. Methods and properties of GstMiningPrep.java class are shown in Figure 4-18.

```
GstMiningPrep mineAll = new GstMiningPrep();
mineAll.mineEvents(gridStations);
```

Figure 4-17 Implementation of mineEvents method call

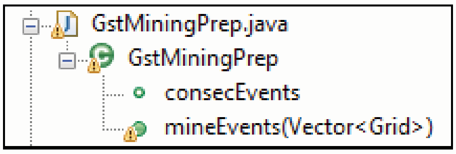


Figure 4-18 Methods and properties of GstMiningPrep.java class

As defined in the previous section only a sample set of stations are processed in this study. Since then, station number of each of the Grid Object in the vector (*gridstations* vector which is sent to the *mineEvents()* method) is controlled by an “if condition”, and if the station number is in the sample set of stations, necessary properties of the station are prepared for the mining.

Similar to the approach given in [9] time windows are defined. A year is divided into 4 time windows and each window is composed of 3 months. Three vectors named month1, month2 and month3 are initiated in the type of EventSetId to hold the eventsets of a time window. Initiation of these vectors is displayed in Figure 4-19.

```
Vector<EventSetId> month1 = new Vector<EventSetId>();
Vector<EventSetId> month2 = new Vector<EventSetId>();
Vector<EventSetId> month3 = new Vector<EventSetId>();
```

Figure 4-19 Initiation of Month Vectors

EventSetId.java class is implemented for defining the records that are inserted into table “EventSetGrid”. Additionally, another property (subsetId) is defined to be used during the creation of subsets. Properties of EventsetId.java class can be seen in Figure 4-20.

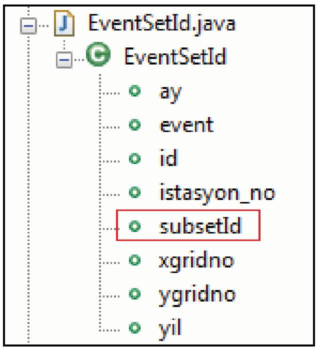


Figure 4-20 Properties of EventsetId.java class

For improving the performance of the implementation and for getting faster results a 16-year sample is chosen, which is from 1975 to 1991.

In order to calculate the frequency of the eventset, each eventset in a month vector, is searched in the same month of the other years. Searching is performed on EventSetGrid table. Records are selected from the “EventSetGrid” table according to the station number, year and month properties. For each of the selected record an object is created from the EvenSetId class. All the properties of grid object are assigned to the properties of this EventSetId object, besides a subsetid is set in the increasing order. All these created objects are added to the month1 vector.

Next step is finding the neighbor stations of this selected station, selecting the eventsets of these neighbor stations in the first month of the time window, creating EventSetId objects for each of them and adding them to the month1 vector.

Since the stations are placed on a grid and each of them are numbered with xgridno and ygridno, finding the neighboring relationship is straight forward. If the xgridno and ygridno differences are smaller than or equal to 1, then it can be said that two stations are neighbors.

All stations are scanned for finding the neighbor stations of the selected station in the second month of the time window to set the elements of month2 vector.

Elements of the month3 vector are found more different than the elements of month2 vector. To find the elements of month3 vector, all the neighbor stations of the selected station, besides all the neighbor stations of the stations selected for month2 vector (neighbors of the neighbors of selected station) are found from EventSetGrid table which occurs in the third window of the time window.

After the vectors are prepared in GstMiningPrep.java class during mining preparation phase, they are ready to be mined. An instance of GstMining.java class is created in GstMiningPrep.java and its *GstMining()* method is called by sending the 3 prepared month vectors and a character for indicating the end of the preparation vectors. Until the year 1991 char ‘a’ is sent to the method; whereas for the year 1991, char ‘f’ is sent. Properties and methods of GstMining.java class are displayed in Figure 4-21.

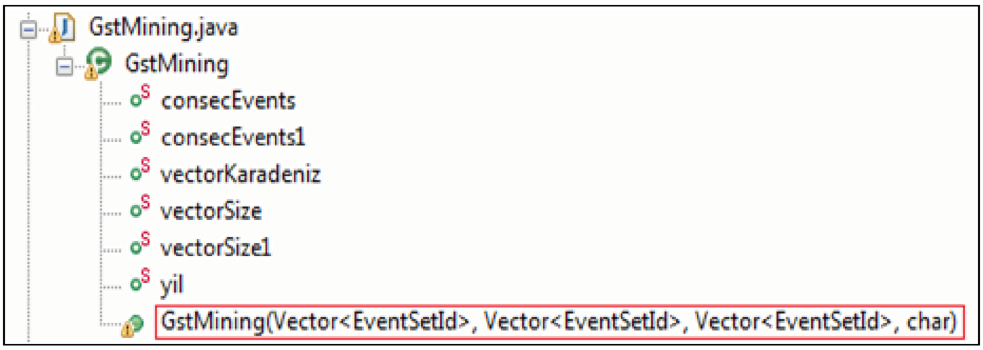


Figure 4-21 Properties and methods of GstMining.java

4.3 Mining and Finding Frequent Rules

Mining and Finding Frequent Rules is the third phase of the implementation. Consecutive eventsets are set and frequencies of the events in those consecutive eventsets are calculated.

Structure to hold the consecutive eventsets and calculation of the frequencies are explained in detail with the implementation point of view in the further sections.

Mining Phase

In GstMining.java class three vectors are initiated named as consecEvents, consecEvents1 and vectorKaradeniz. consecEvents and consecEvents1 vectors are created in the type of ConsecEvents.java class, vectorKaradeniz is created in the type of ConsecEventsPercentage.java class. Initiation of Consecutive events vectors is shown in Figure 4-22.

```
public static Vector<ConsecEvents> consecEvents = new Vector<ConsecEvents>() ;
public static Vector<ConsecEvents> consecEvents1 = new Vector<ConsecEvents>() ;
public static Vector<ConsecEventsPercentage> vectorKaradeniz = new Vector<ConsecEventsPercentage>() ;
```

Figure 4-22 Initiation of Consecutive events vectors

ConsecEvents and ConsecEventsPercentage classes have the same properties except the percentage parameter. This parameter which is defined in the type of float is used for the calculation of the frequency of the output rules. Consec1 and Consec2 vectors defined in this class hold elements in the type of EventSetId.java class. Difference of ConsecEvents.java and ConsecEventsPercentage.java classes are displayed in Figure 4-23.

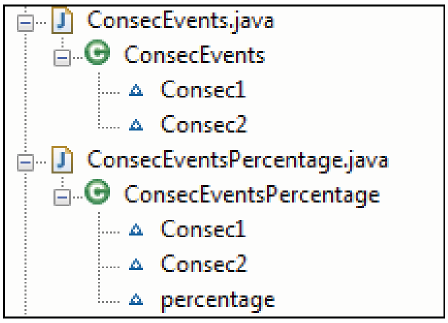


Figure 4-23 Properties of ConsecEvents.java and ConsecEventsPercentage.java

Three vectors named subsetId1, subsetId2 and subsetId3 are initiated which hold the subsetIds of the evensets of month1, month2 and month3 vectors.

For ease of powerset creation, integer subset numbers have been assigned to the elements of the eventsets that are created for time windows in the preprocessing phase. All the subsets of eventsets are calculated. Next, by the help of power set property, all of the subsets of eventsets of month1 and month2, besides month2 and month3 are calculated.

VectorSize.java is written for determining of the size of the eventsets per year. Properties of this class are displayed in Figure 4-24.

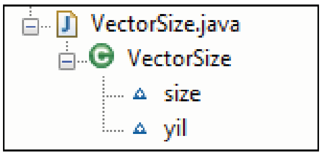


Figure 4-24 Properties of VectorSize.java

FindFrequentRules.java

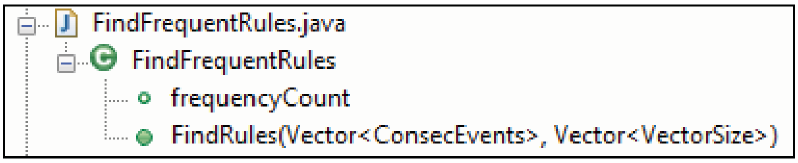


Figure 4-25 Properties and methods of FindFrequentRules.java

Once the flag is 'f' which indicates that they are the last eventsets (month vectors) that are sent to GstMining.java class, prepared ConsecEvents and ConsecEvents1 vectors are ready to be sent to the *FindRules()* method of FindFrequentRules.java class for finding the frequency of the rules. Properties and methods of FindFreuquentRules.java class ara displayed in Figure 4-25.

An instance of FindFrequentRules.java class is created. Additionally, 2 new vectors named as resultSet and resultSet1 are initiated which hold elements in the type of ConsecEventsPercentage. Initiation of resultSet vectors are displayed in Figure 4-26.

```
FindFrequentRules findRules = new FindFrequentRules();
Vector<ConsecEventsPercentage> resultSet  = new Vector();
Vector<ConsecEventsPercentage> resultSet1  = new Vector();
```

Figure 4-26 Initiation of ResultSet Vectors

Prepared consecEvents and vectorSize vectors are sent to the *FindRules()* method of FindFrequentrules.java class. *FindRules()* method returns a vector which holds elements in the type of ConsecEventsPercentage. Returned vectors of this method are assigned to resultSet and resultSet1 vectors. This assignment is shown in Figure 4-27.

```
resultSet = findRules.FindRules(consecEvents,vectorSize);
resultSet1 = findRules.FindRules(consecEvents1,vectorSize1);
```

Figure 4-27 Returned Resultset vectors by FindRules() method

Frequency count is set to 10. Since there are 16 years to look for frequent eventsets, if an eventset occurs in at least 10 years out of 16, it is assumed that this eventset is frequent.

ConsecEvents and ConsecEvents1 vectors hold the whole eventsets for 16 years. vectorsSize and vectorSize1 vectors include the information of the size of the eventsets for each year. This is a performance issue to determine the ranges and sizes of the eventsets for each year.

Then for each of the eventsets in the restricted set (eventsets of a year), eventset's Consec1 vector is compared with the Consec1 vectors of the other eventsets in the other years. Station numbers and events are compared to determine if the eventsets are identical or not. If they are, then a boolean parameter is set to true, and if that parameter is true(which indicates that consec1 vectors are equal) then Consec2 vector of the eventset is compared to the Consec2 vectors of the other eventsets in the other years. This is another issue for increasing the performance of the implementation. If Consec1 vectors are not equal, it is not necessary to compare the Consec2 vectors. And if for each of the eventset that has the identical Consec1 and Consec2 vectors with the one we selected, the frequency count is increased by one.

And if the frequency count of the selected eventset is greater than or equal to the frequency count, then that eventset is frequent. A new object is created in the type of ConsecEventsPercentage class. Consec1 and Consec2 vectors of the frequent eventset are assigned to the Consec1 and Consec2 vectors of this recently created object. Also percentage is found by dividing the frequency count to the number of years (16 for this implementation). And calculated percentage is assigned to the percentage property of the created object. Finally this object is added to the resultSet vector which is created for holding the frequent eventsets.

After all the frequent eventsets are found and all those found frequent eventsets are added to the resultSet vector, a new object is created from the PruningResultSet.java class. resultSet vector which hold the elements in the type of ConsecEventsPercentage, is sent to the *Pruning()* method of this object. Return statement of *GstMining()* method and the Pruning() method call is displayed in Figure 4-28.

```
PruningResultSet prunSet = new PruningResultSet();
return prunSet.Pruning(resultSet);
```

Figure 4-28 Return statement of GstMining method

And the result of the pruning process is returned from FindFrequentRules.java class to the GstMining.java class where the *FindRules()* method of FindFrequentRules class is called from.

PruningResultSet.java

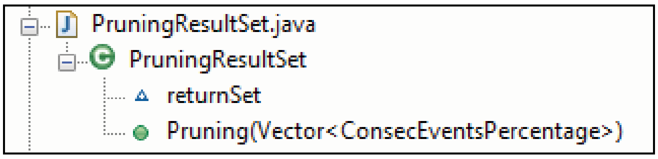


Figure 4-29 Properties and Methods of PruningResultSet.java class

PruningResultset.java class is implemented in order to eliminate the subsets of the frequent eventsets. If an eventset is frequent all its subsets are frequent either. In GstMining.java class all the subsets of the eventsets are found and added separately to the ConsecEvents and ConsecEvents1 vectors. These vectors are sent to the *FindRules()* method.

In the Pruning method each of the frequent eventsets in the sent vector is compared with the other eventsets in the same vector. Consec1 and Consec2 vectors of each eventset are compared with the Consec1 andConsec2 vectors of other eventsets with their station number, event and month properties. If all of them are the same, they are the subsets of the eventset and they are pruned from the resultset and added to a new “returnSet” vector. After all the eventsets are controlled and compared with each other and returned set is created with the pruned eventsets, this set is returned to the GstMining.java class as a result of FindRules method of FindFrequentRules.java class.

Once all the frequent eventsets are found and returned to the GstMining.java class, Output files are written. While writing the frequent eventsets to the output files, each frequent eventset is added to the vectorKaradeniz. At the end of the writing process, vectorKaradeniz includes all the frequent eventsets for all of the stations and for all of the years. Then an instance of OverMining.java is created in GstMiningPrep.java class and vectorKaradeniz is sent to the *mine()* method of this class throughout GstMining object’s static parameter. Related method call is displayed in Figure 4-30.

```
GstMining mine = new GstMining();
OverMining overmine = new OverMining();
overmine.overMine(mine.vectorKaradeniz);
```

Figure 4-30 Overmining method call

4.4 Mining Rules

'Mining Rules' is the last phase of the implementation. OverMining.java class is implemented to find new rules from the generated rules. It is the generalization of the rules. Properties and methods of OverMining.java class are displayed in Figure 4-31.

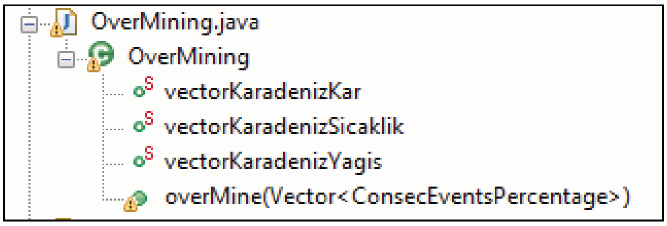


Figure 4-31 Properties and Methods of OverMining.java class

Three new vectors are initiated which hold the elements in the type of ConsecEventsPercentage. Initiation of event type vectors is displayed in Figure 4-32.

```
public static Vector<ConsecEventsPercentage> vectorKaradenizYagis = new Vector<ConsecEventsPercentage>() ;
public static Vector<ConsecEventsPercentage> vectorKaradenizKar = new Vector<ConsecEventsPercentage>() ;
public static Vector<ConsecEventsPercentage> vectorKaradenizSicaklik = new Vector<ConsecEventsPercentage>() ;
```

Figure 4-32 Initiation of Overmining Vectors

vectorKaradeniz contains all of the evensets. Event characters in the Consec1 and Consec2 vectors of those eventsets are controlled and vectorKaradeniz is separated into 3 vectors named vectorKaradenizYagis, vectorKaradenizKar and vectorKaradenizSicaklik according to the event types.

For each event type 4 situations are defined for representing event changes due to x grid number changes and 4 situations are defined for representing event changes due to y grid number changes. Variables are initiated to count the number of the elements for each defined situation. Initiation of situation variables due to ygridno change is displayed in Figure 4-33, and Initiation of situation variables due to xgridno change is displayed in Figure 4-34.

```
int yagisGridUpEventUpCount = 0;
int yagisGridDownEventDownCount = 0;
int yagisGridUpEventDownCount = 0;
int yagisGridDownEventUpCount = 0;

int karGridUpEventUpCount = 0;
int karGridDownEventDownCount = 0;
int karGridUpEventDownCount = 0;
int karGridDownEventUpCount = 0;

int sicaklikGridUpEventUpCount = 0;
int sicaklikGridDownEventDownCount = 0;
int sicaklikGridUpEventDownCount = 0;
int sicaklikGridDownEventUpCount = 0;
```

Figure 4-33 Initiation of Count Variables for y grid no changes

```
int yagisGridUpEventUpCountxgridno = 0;
int yagisGridDownEventDownCountxgridno = 0;
int yagisGridUpEventDownCountxgridno = 0;
int yagisGridDownEventUpCountxgridno = 0;

int karGridUpEventUpCountxgridno = 0;
int karGridDownEventDownCountxgridno = 0;
int karGridUpEventDownCountxgridno = 0;
int karGridDownEventUpCountxgridno = 0;

int sicaklikGridUpEventUpCountxgridno = 0;
int sicaklikGridDownEventDownCountxgridno = 0;
int sicaklikGridUpEventDownCountxgridno = 0;
int sicaklikGridDownEventUpCountxgridno = 0;
```

Figure 4-34 Initiation of Count Variables for x grid no changes

Characters which represent the events are converted into numeric values from 1 to 6 according to their intensities in the increasing order. Within the frequent consecutive eventsets, numeric values of the events of Consec1 eventsets are compared with the numeric values of the events of Consec2 eventsets. Grid numbers of the stations in the Consec1 and Consec2 eventsets are also compared. Situations are controlled with ‘if conditions’ and related variable count is increased. After all the eventsets are controlled, and variable counts are determined, values of the parameters and short descriptions are written to the output file named “EventChanges”.

CHAPTER 5

INTERPRETATION OF THE GENERATED SPATIO-TEMPORAL PATTERNS & EVALUATION OF THE RESULTS

This section contains the interpretation of the generated spatio-temporal patterns. Sample output files are selected to explain the details of the file contents and the way of interpretation. Section also contains the interpretation of generated higher-level rules with "Mining Rules" phase. Experiments that are applied on the data set to monitor the change of the events for different temporal intervals are also explained in this section.

5.1 Interpretation of the Generated Spatio-Temporal Patterns

In the experiments presented in this section is performed by application of the developed technique on the Meteorological Data set (as described in Proposed Approach section). Maximum 20 output files can be generated for a station. However, if there are not any frequent eventsets for some of the months, output files are not written for those particular months and number of the output files may decrease. A sample set of output files that are written for station 17022 (Station Name: Zonguldak) are given in the appendices.

As it was explained in the previous sections, time windows which are used for finding temporal associations are composed of 3 months, and extracted flow patterns indicate the change of events in consecutive 3 months within the same time window. Two output files are written for each time window. Samples from the written output files will be used for illustration.

In the sample set, there are 20 output files. Number of output files indicates that frequent eventsets have been found for every month. Output files can be grouped for each time window. Output files 17022— month —1 and 17022—1— month —2 include the generated patterns of the first time window. Output files are composed of the eventsets which have occurred consecutively in January, February and March. Since the numeric representation is used for months, "month —1" in the name of the file indicates that file contains the eventsets of the first month which is January and the consecutive month which is February. "1— month —2" in the name of the second output file indicates that file includes the eventsets of the time window which starts with the first month, but it is the second file for that time window. It also indicates that the file contains the eventsets of the second month which is February and the consecutive of second month which is the third month March. Files 17022— month —1 and 17022—1— month —2 are written for January, February and

March. Likewise files 17022—ay—2 and 17022—2— month —3 are written for February, March and April. Similarly all the output files are written for every time window if they contain frequent eventsets. It is the result of sliding method that is used specifically for this study. As already mentioned in the previous sections, by this method the loss of any association between different time windows is prevented. Output file 17022— month —1 will be used for identifying the contents of the files. Content of this output file is shown in Figure 5-35.

17022-- month —1

```
**************************** Percentage 62.5
month: 1 17070  event is a
month: 2 17070  event is 2
month: 2 17070  event is a
**************************** Percentage 62.5
month: 1 17022  event is b
month: 1 17070  event is a
month: 2 17070  event is 2
**************************** Percentage 62.5
month: 1 17022  event is b
month: 1 17070  event is a
month: 2 17612  event is b
**************************** Percentage 62.5
month: 1 17022  event is b
month: 1 17070  event is a
month: 1 17602  event is b
month: 2 17070  event is a
**************************** Percentage 62.5
month: 1 17022  event is 6
month: 1 17070  event is a
month: 2 17070  event is a
```

Figure 5-35 Content of the Output Files

Each frequent eventset found is displayed with its frequency at the beginning of the eventset. Percentage of frequency is calculated. Since a 16-year period is taken to search for the frequent events, %62.5 means that eventset has been occurred in 10 years out of 16 years. Already %62.5 is the support (minimum threshold) that is chosen for finding the frequent eventsets. In order to facilitate the interpretation of experiments, the event classification tables are presented in Figure 5-36.

Precipitation	Event
precipitation <= 25.0 mm	1
25.0 mm <= precipitation <=50.0 mm	2
50.0 mm <= precipitation <=75.0 mm	3
75.0 mm <= precipitation <=100.0 mm	4
100.0 mm <= precipitation <=125.0 mm	5
precipitation >= 125.0 mm	6

Number of Snowy Days	Event
snowyDays <= 5	g
5 <= snowyDays <=9	h
9 <= snowyDays <=13	i
13 <= snowyDays <=17	j
17 <= snowyDays <=21	k
snowyDays >= 21	l

Average Temperature	Event
AvgTemp <= 5.0 C°	a
5.0 C°<= AvgTemp <=10.0 C°	b
10.0 C°<= AvgTemp <=15.0 C°	c
15.0 C°<= AvgTemp <=20.0 C°	d
20.0 C°<= AvgTemp <=25.0 C°	e
AvgTemp >= 25.0 C°	f

Figure 5-36 Event classification Tables

Frequent eventsets consist of at least 2 eventsets. But there is not a restriction for the maximum number of eventsets within a frequent eventset. The month of the eventset which the event has been occurred, the station number and the character for representing the occurring event are written for each frequent eventset. It should be kept in the mind that stations that are written in the same eventset should have a neighborhood relationship (being neighbors or neighbor of neighbors).

```
***************************** Percentage 62.5
month: 1 17022  event is b
month: 1 17070  event is a
month: 2 17070  event is 2
```

Figure 5-37 Frequent Eventset from Output file "17022— month —1"

The pattern given in Figure 5-37 corresponds to the following description in natural language: *"In January when average temperature is between 5.0 C° and 10.0 C° in Zonguldak (17022) and average temperature is less than 5.0 C° in Bolu (17070), in February precipitation is between 25.0 and 50.0 mm in Bolu (17070)".* The support of this pattern is 62.5% (It occurs 10 times out of 16 years). Obviously, not all of the frequent eventsets are interesting; however, it is important to capture the general flowing patterns.

Frequent eventsets (they can also be called as '"frequent association rules") found for the same time window but written into different output files are spatially and temporally associated. They are spatially associated because stations written in the second output file of the time window are the neighbor of the selected station or the neighbor of the selected station's neighbors. And they are temporally associated because they have occurred in the consecutive months. In spite of these associations, it is not possible to relate the 2 eventsets that include the same stations and event types in the consecutive months. However it is possible to observe the general behavior of the event changes.

Sample eventsets (rules) that are selected from different output files are interpreted.

```
17030–5--month—6 (name of the output file which the frequent eventset is taken from)
***************************** Percentage: 81.25
month: 6 17085  event is e
month: 6 17622  event is g
month: 7 17030  event is e
month: 7 17681  event is 1
```

Figure 5-38 Frequent Eventset from Output file “17030--5--month --6”

Consecutive eventset is displayed in Figure 5-38. Average temperature is between 20.0 C° and 25.0 C° in June in Amasya (17085). Besides, average temperature is between 20.0 C° and 25.0 C° in July in Samsun (17030) and precipitation is less than 25.0mm in July in Zile (17681). This composite eventset has occurred 13 times out of 16 (support of the rule is %81.25). High average temperature in Amasya in June resulted minimum precipitation (event ‘1’ stands for the minimum precipitation) in Zile in July. Temperature event may not be the causative event for the precipitation in the neighbor location but this is a flowing pattern and the sequence of the events.

Another sample is from the output file written for Samsun. It can be noticed that number of the elements in the frequent eventsets may vary for different consecutive eventsets. A sample part from the content of the output file “17030—7—month--8” is shown in Figure 5-39.

```
17030–7--month--8
***************************** Percentage: 68.75
month: 8 17085  event is 1
month: 8 17622  event is g
month: 8 17622  event is e
month: 9 17681  event is g
month: 9 17681  event is 1
month: 9 17681  event is d
```

Figure 5-39 Frequent Eventset from Output file “17030--7--month --8”

When precipitation is less than 25.0mm in August in Amasya (17085) and average temperature is between 20.0 C° and 25.0 C° in August in Bafra (17622), precipitation is less than 25.0mm in September in Zile (17681) and average temperature is between 15.0 C° and 20.0 C° in September in Zile (17681). Events related to snowy days (‘g’) are not considered, because it is obviously normal that number of snowy days are smaller than 5 in August and September in Black Sea Region.

```
17030--10--month--11
**************************** Percentage: 62.5
month: 11 17085  event is b
month: 11 17622  event is g
month: 12 17030  event is g
month: 12 17083  event is a
month: 12 17681  event is a
```

Figure 5-40 Frequent Eventset from Output file “17030--10--month --11”

Consecutive eventset given in Figure 5-40 indicates that when the average temperature is between 5.0 C° and 10.0 C° in November in Amasya (17085), average temperature is less than 5.0 C° in Merzifon (17083) and in Zile (17681) in December. This eventset has been observed in 10 years out of 16.

```
17030--month--5
**************************** Percentage: 62.5
month: 5 17085  event is 2
month: 5 17622  event is g
month: 6 17085  event is e
month: 6 17622  event is g
```

Figure 5-41 Frequent Eventset from Output file “17030--month --5”

It is also possible to observe the flow of events at the same location. Consecutive eventset given in Figure 5-41 represents the change of events in Amasya (17085) between May and June. While precipitation is between 25.0 and 50.0 mm in Amasya in May, average temperature is between 20.0 C° and 25.0 C° at the same location in June.

Most of the frequent evensets include the event ‘g’ which indicates that number of the snowy days is smaller than 5 in a month. It is normal that coast of the Black Sea Region is not so snowy. For Example, if the sample set of stations was chosen from East Anatolia, eventsets would have included different snow events.

Although the patterns generated are not so similar to the ones explained in [9], these are the results that can be generated with this meteorology data set. Events are spatially and temporally related, but it may not be correct to relate the events in a causative manner.

```
17033--5--month--6
**************************** Percentage: 75.0
month: 6 17624  event is g
month: 6 17624  event is d
month: 7 17033  event is e
```

Figure 5-42 Frequent Eventset from Output file “17033—5--month --6”

According to the consecutive eventset given in Figure 5-42, high temperature in Ünye (17624) in June depicts that the average temperature in Ordu (17033) in July is between 20.0 C° and 25.0 C°. Support of the generated pattern is %75.

```
17033--month--8
***************************** Percentage: 68.75
month: 8 17033  event is e
month: 8 17034  event is e
month: 9 17034  event is d
month: 9 17624  event is d
```

Figure 5-43 Frequent Eventset from Output file "17033--month --8"

It is possible to observe the decrease at the average temperature in the evenset displayed in Figure 5-43. In Ordu (17033) and in Giresun (17034) average temperature is between 20.0 C° and 25.0 C° in August. But with the change of the season from summer to autumn (change of the month from August to September) average temperature decreases by 5 C° in Giresun and in Ünye (17624).

```
17034--1--month--2
***************************** Percentage: 68.75
month: 2 17033  event is b
month: 3 17034  event is b
```

Figure 5-44 Frequent Eventset from Output file "17034—1--month --2"

Generated flow pattern that is displayed in Figure 5-44 may be interpreted in a different way. Average temperature in Ordu (17033) in February is approximately same with the average temperature observed in Giresun in March. Increase in the temperature is expected with the change of the month from February to March. Since both stations have approximately the same temperature in different months, one may comment that weather of Ordu is warmer than the weather in Giresun. Obviously many different parameters should be considered for such a decision.

```
17034-- month --6
***************************** Percentage: 68.75
month: 6 17033  event is d
month: 7 17033  event is e
```

Figure 5-45 Frequent Eventset from Output file "17034--month --6"

Due to the consecutive eventset given in Figure 5-45, approximately 5 C° of increase in the average temperature is observed in Ordu (17033) with the change of the month from June to

July. Since average temperature is classified in 5 C° intervals, this comment can be made as the event is flown from d to e at the same location.

```
17037-- month --1
**************************** Percentage: 68.75
month: 1 17088  event is 1
month: 1 17626  event is b
month: 2 17088  event is a
**************************** Percentage: 68.75
month: 1 17088  event is a
month: 1 17626  event is b
month: 2 17088  event is 2
month: 2 17088  event is a
```

Figure 5-46 Frequent Eventset from Output file “17037--month --1”

Some of the frequent patterns in the output files can be interpreted together. In the consecutive eventsets displayed in Figure 5-46, different properties of Gümüşhane (17088) appear in different frequent eventsets of the same output file. In the first set, precipitation event of Gümüşhane has been investigated. And in the second set, average temperature of the same station has been considered. It is also possible to observe the event changes of the stations by searching different patterns in same and different output files. The increase in the precipitation of 17088 can be observed by referring to different frequent eventsets. It is less than 25 mm in January (event is ‘1’), but it is between 25 and 50 mm in February (event is ‘2’).

Consecutive eventsets listed in Figure 5-47 show that precipitation is more than 125mm (event is ‘6’) in Rize (17040) and in Hopa (17042) in most of the months of the year (from May to December).

```
17040--5--month --6

**************************** Percentage: 81.25
month: 6 17042  event is 6
month: 7 17040  event is g
month: 7 17040  event is e
month: 7 17668  event is g

17040--7--month --8

**************************** Percentage: 81.25
month: 8 17042  event is e
month: 9 17040  event is g
month: 9 17040  event is 6
month: 9 17045  event is d

17040--8--month --9

**************************** Percentage: 68.75
month: 9 17042  event is 6
month: 9 17042  event is d
month: 10 17040  event is g
month: 10 17040  event is 6
month: 10 17045  event is c
```

Figure 5-47 Frequent Eventsets from different Output files of station 17040

Another comment that can be made on these eventsets is about the average temperature. Even in winter, average temperature is not under 5 C° (In November in Rize (17040) average temperature is between 10.0 C° and 15.0 C°).

```
17040--9--month --10

**************************** Percentage: 75.0
month: 10 17042  event is 6
month: 11 17040  event is c
**************************** Percentage: 68.75
month: 10 17042  event is 6
month: 11 17040  event is g
month: 11 17045  event is b


17040--10--month --11

**************************** Percentage: 68.75
month: 11 17042  event is 6
month: 12 17040  event is 6
month: 12 17668  event is 1
month: 12 17668  event is a
```

Figure 5-48 Frequent Eventsets from different Output files of station 17040 (cont.)

Frequent Consecutive eventset displayed in Figure 5-48 indicates that Hopa (17042) and Artvin (17040) are rainy in November and in December, but their neighbor Oltu (17668), which is actually in East Anatolia, has a harsh climate in winter season. Average temperature is less than 5 C° and precipitation is less than 25mm in Oltu in December. This indicates that the actual precipitation is generally in the form of snow rather than rain due to cold weather.

```
17040--month --3
**************************** Percentage: 93.75
month: 3 17040  event is b
month: 4 17042  event is c
```

Figure 5-49 Frequent Eventset from Output file “17040--month --3”

Some of the frequent patterns have significant support values (percentage of the frequencies) as displayed in Figure 5-49.

There are only 2 frequent eventsets in the output file (`17045-- month --2`) shown in Figure 5-50. Both of the eventsets include the average temperature event that occurred in Oltu (17668) in February. But second eventsets in those eventsets are different. And it can be said that those second eventsets are not associated with each other although they are associated with the same eventset (first eventset in these eventsets).

```
17045--month --2
***************************** Percentage: 62.5
month: 2 17668  event is a
month: 3 17042  event is b
***************************** Percentage: 62.5
month: 2 17668  event is a
month: 3 17668  event is 1
```

Figure 5-50 Frequent Eventsets from Output file “17045--month --2”

```
17045--month --5
***************************** Percentage: 62.5
month: 5 17045  event is d
month: 6 17042  event is 6

17045--5--month --6
***************************** Percentage: 68.75
month: 6 17042  event is 6
month: 6 17668  event is g
month: 6 17668  event is d
month: 7 17666  event is e
```

Figure 5-51 Frequent Eventsets from different Output files of station 17045

Since frequent eventsets of the consecutive 3 months are discovered in common, it might be meaningful to interpret the rules of the consecutive output files together. Although they are not directly related, they might show the flow of the events in the correct order. Consecutive eventsets shown in Figure 5-51 are selected from the consecutive output files which are '17045-- month --5' and '17045--5-- month –6'. Flow of the events that occurred in the neighbor stations from May to July can be observed in these eventsets, Events occurred in the fifth and seventh months can not be associated directly. That is, average temperature event'd' in Artvin (17045) is not directly related to the average temperature event 'e' observed in Ispir (17666) in July. But it is obvious that; these consecutive output files include eventsets which are spatially and temporally related.

Eventsets in different output files can be interpreted consecutively as the following; Average temperature is between 15.0 C° and 20.0 C° in Artvin (17045) in May and precipitation is more than 125.0 mm in Hopa (17042) in June. Sequentially average temperature is between 15.0 C° and 20.0 C° in Oltu (17668) in June and average temperature is between 20.0 C° and 25.0 C° in Ispir (17666) in July.

```
17045-- month --8
**************************** Percentage: 62.5
month: 8 17042  event is e
month: 8 17668  event is g
month: 8 17668  event is e
month: 9 17042  event is 6


17045--8-- month --9
**************************** Percentage: 68.75
month: 9 17042  event is 6
month: 10 17045  event is c
**************************** Percentage: 68.75
month: 9 17042  event is d
month: 10 17045  event is c
```

Figure 5-52 Frequent Eventsets from different Output files of station 17045

Consecutive eventsets displayed in Figure 5-52 which exist in different output files can be interpreted similarly. Average temperature event 'e' occurred in Hopa (17042) and in Oltu (17668) in August is flown to the precipitation event '6' in Hopa in September in the first output file (`17045-- month --8`). And in the second output file, in the first eventset it is observed that precipitation event '6' which occurred in Hopa in September is flown to the temperature event 'c' which occurred in Artvin in October. Since there is a common event in both of the eventsets of different output files, these events may be interpreted together consecutively. But the second eventset in the second output file (`17045--8-- month --9`) can not be associated with the composite eventset mentioned above. Because this eventset does not include a common event with the evenset in the first output file.

```
17088--month --10
**************************** Percentage: 68.75
month: 10 17088  event is c
month: 10 17037  event is d
month: 11 17037  event is c

17088--10--month --11
**************************** Percentage: 75.0
month: 11 17037  event is c
month: 12 17088  event is a
month: 12 17626  event is b
```

Figure 5-53 Frequent Eventsets from different Output files of station 17088

Same kind of events (average temperature event) occurred in different stations in different months give us a clue while comparing the climates of the cities. Due to the consecutive eventsets given in Figure 5-53, while average temperature is between 10.0 C° and 15.0 C° in Gümüşhane (17088) in October, it is between 15.0 C° and 20.0 C° in Trabzon (17037). Also, while average temperature is less than 5.0 C° in Gümüşhane (17088) in December, it is between 5.0 C° and 10.0 C° in Akçaabat (17626).

```
17626--7--month --8
**************************** Percentage: 62.5
month: 8 17037  event is e
month: 9 17088  event is 1
month: 9 17088  event is d
month: 9 17626  event is d
```

Figure 5-54 Frequent Eventset from Output file “17626—7--month --8”

Another observation that can be made on the generated spatio-temporal patterns is about the climate. Considering only one of the events is not sufficient for defining the climatological properties. But in the consecutive eventset displayed in Figure 5-54, 2 different properties of the same station are given for the same month. Precipitation is less than 25 mm and average temperature is between 15.0 C° and 20.0 C° in Gümüşhane (17088) in September. It is obvious that precipitation is much less in Gümüşhane when it is compared to the precipitation of other cities in Black Sea Region. Weather is warm and arid. Location of the city is the main reason for this climatic change. It is positioned between Black Sea Region and East Anatolia. Since then this city has taken the combinational climatological properties of these two regions.

5.2 Interpretation of Generated Higher-level Rules with 'Mining Rules' Phase

Following 'Event Changes' file is the output file of the 'Mining Rules' Phase. Content of this file is directly related to the input parameters. It will be beneficiary to remind the concept of 'Mining Rules' phase. All the frequent eventsets generated for the input stations are collected in a vector named 'VectorKaradeniz'. All eventsets are separated into three different vectors named as vectorKaradenizYagis, vectorKaradenizKar and vectorKaradenizSicaklik due to the event type, and all these eventsets are examined by comparing the intensity of the events and comparing the x and y grid numbers. But it should be kept in mind that compared eventsets have a neighboring relationship. For example, the intensities of the precipitation events in Bolu and in Artvin are not compared, and a rule for indicating the change of event between further locations can not be generated. Instead, events in the neighbor stations are compared and considered.

--EventChanges—

```
********Event Changes***************
Size of Vector Karadeniz Temperature :639
Size of Vector Karadeniz Precipitation :56
Size of Vector Karadeniz Snowy Days :679

******event changes due to the ygridno change******

sicaklikGridUpEventUpCount = 38 ygridno artinca sicaklik artmistir
sicaklikGridDownEventDownCount = 49 ygridno azalinca sicaklik
azalmistir
sicaklikGridUpEventDownCount = 57 ygridno artinca sicaklik
azalmistir
sicaklikGridDownEventUpCount = 132 ygridno azalinca sicaklik
artmistir

*****event changes due to the xgridno change*****

sicaklikGridUpEventUpCountxgridno = 123 xgridno artinca sicaklik
artmistir
sicaklikGridDownEventDownCountxgridno = 60 xgridno azalinca sicaklik
azalmistir
sicaklikGridUpEventDownCountxgridno = 50 xgridno artinca sicaklik
azalmistir
sicaklikGridDownEventUpCountxgridno = 8 xgridno azalinca sicaklik
artmistir
```

Figure 5-55 “Event Changes” Output file

A sample set of stations which are mostly located on the coast of Black Sea Region has been chosen. Since these stations are not so snowy, frequent snow events included in the frequent eventsets do not differentiate from station to station. Number of snowy days are always smaller than 5 and the related event is ‘g’ in the frequent eventsets. Because of that, when the intensities of snow events are compared, no difference could be found. As a result, Event Changes file that is displayed in Figure 5-55 does not include any outputs related to snow events.

When the values of the variables in the output file are inspected, average temperature change due to x and y grid number changes can be interpreted. ‘sicaklikGridDownEventUpCount’ variable has the biggest value in ‘event changes due to the ygridno change’ section. And ‘sicaklikGridUpEventUpCountxgridno’ variable has the biggest value in ‘event changes due to the xgridno change’ section.

Average temperature of a station with a greater y grid number is higher than the average temperature of another station with a smaller y grid number. In other words, stations located in the East generally have higher average temperatures than of the ones located to the West of them. This rule has a %47 support value (i.e. it has been observed 132 times out of 276).

When the values of the parameters that are written for the event changes due to the x grid number changes are analyzed, it is observed that average temperature rises with the increase in the x grid number. This rule has %51 support value (i.e. it has been observed 123 times out of 241). It might be said that, generally, stations located up North have higher average temperatures compared to the average temperature of the stations located to the South of them.

When the general climatological characteristics of Black Sea Region are considered, straightness of the output rules can be proved. Inner Black Sea Region shows the properties of Terrestrial climate. From this point of view, it is normal to have higher average temperature on the northern side (along the coast of Black Sea Region). And it is known that, East side of the Black Sea Region which is called as ‘East Black Sea’ has a warmer and more rainy climate when it is compared to the climate of the ‘Middle Black Sea’. Generated rule which indicates that stations on the eastern side have higher average temperatures verifies the general characteristics of the Blacksea climate.

5.3 Experiments on the Climate Change Analysis

Proposed approach can be used to analyze the available data set with many different aspects. Temporal or spatial definitions might be changed to compare the generated patterns. In this study a sample set of stations is chosen to generate the frequent patterns. Changing the sample set will affect the generated results from the spatial point of view. In addition changing the value of neighboring relationship will change the spatial associations, and will generate different patterns.

Another experiment is conducted on the analysis of climate change over the temporal periods. Three different temporal sets are created for the years 1986 to 1993, 1993 to 2000 and 2000 to 2007. Patterns and generalized rules generated from these temporal sets are compared. Output files which are written at the end of “Mining Rules” phase are interpreted.

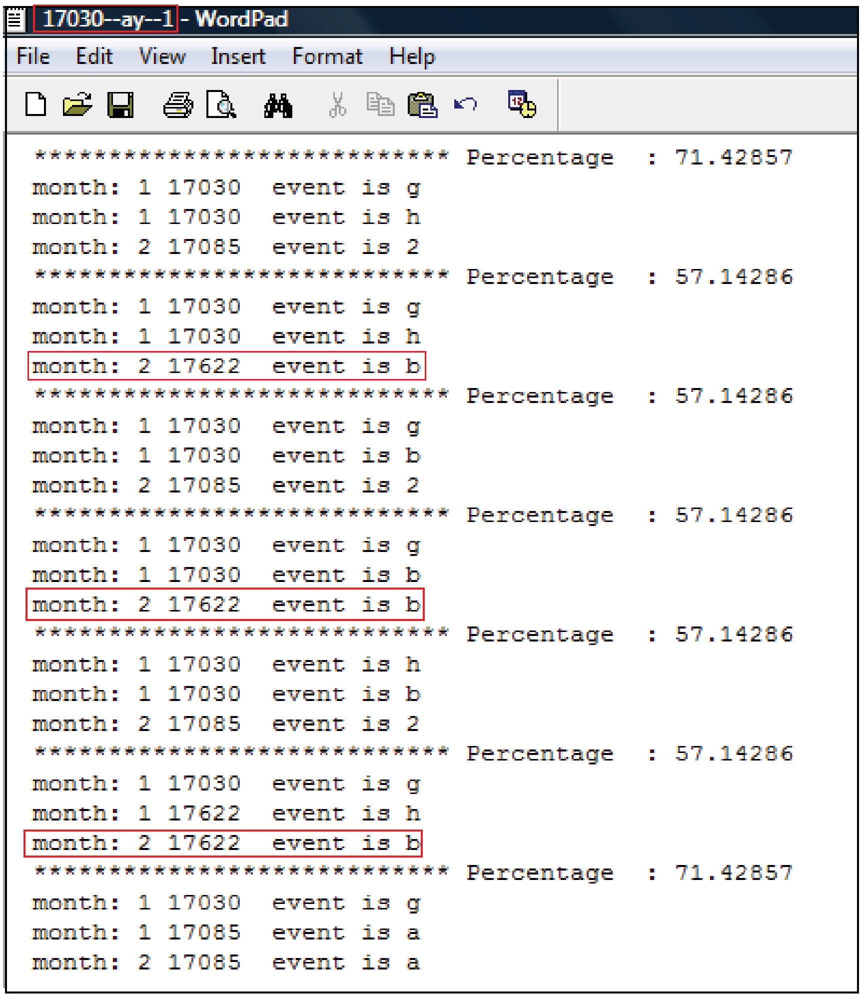
As a difference, for this experiment, during the classification of snow events, one more event is defined to differentiate the case which shows that there are not any snowy days in a month. Updated snow event table is displayed in Table 5-16.

Table 5-16 Updated Snow Event Table

Number of Snowy Days	Event
snowyDays = 0	g
0 < snowyDays <= 4	h
4 <= snowyDays <= 7	i
7 <= snowyDays <= 10	j
10 <= snowyDays <= 13	k
13 <= snowyDays <= 17	l
snowyDays >= 17	m

First section for this experiment was the creation of the output files for the defined temporal sets. Necessary parameters which specify the temporal inputs and frequency count are changed. After getting the output files for the defined sets, it was necessary to compare them to determine the affects of the temporal changes on the events. It was obvious that manual analysis of the output files was hard and ineffective. Implementation is improved to compare the output files automatically and to document the found differences.

Output files that are written to display the frequent eventsets might include the same eventset in many different consecutive eventsets. This situation obstructs the comparison of the files. All the frequent eventsets of an output file are listed in a different output file. Those output files are entitled by the addition of “output” front of the name of the output file they are generated from. Output file “17030—ay—1” is given as an example in Figure 5-56. Original output file includes the frequent eventsets which exist in many different consecutive eventsets. Generated output file "output17030—ay--1" displayed in Figure 5-57 contains only the list of the frequent eventsets.

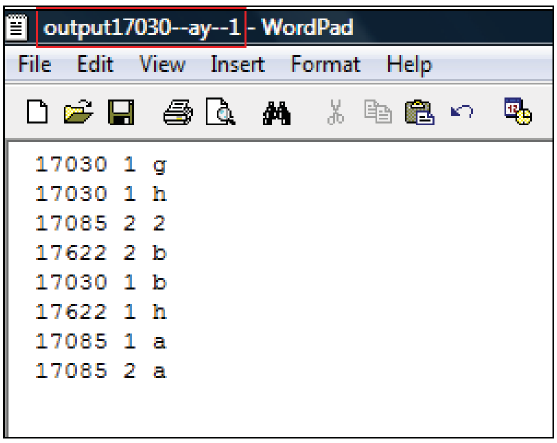


17030--ay--1 - WordPad

File Edit View Insert Format Help

```
***************************** Percentage  : 71.42857
month: 1 17030  event is g
month: 1 17030  event is h
month: 2 17085  event is 2
***************************** Percentage  : 57.14286
month: 1 17030  event is g
month: 1 17030  event is h
month: 2 17622  event is b
***************************** Percentage  : 57.14286
month: 1 17030  event is g
month: 1 17030  event is b
month: 2 17085  event is 2
***************************** Percentage  : 57.14286
month: 1 17030  event is g
month: 1 17030  event is b
month: 2 17622  event is b
***************************** Percentage  : 57.14286
month: 1 17030  event is h
month: 1 17030  event is b
month: 2 17085  event is 2
***************************** Percentage  : 57.14286
month: 1 17030  event is g
month: 1 17622  event is h
month: 2 17622  event is b
***************************** Percentage  : 71.42857
month: 1 17030  event is g
month: 1 17085  event is a
month: 2 17085  event is a
```

Figure 5-56 Existence of frequent eventsets in different consecutive eventsets



output17030--ay--1 - WordPad

File Edit View Insert Format Help

```
17030 1 g
17030 1 h
17085 2 2
17622 2 b
17030 1 b
17622 1 h
17085 1 a
17085 2 a
```

Figure 5-57 Generated output files contains the list of the frequent eventsets

After the frequent eventsets are listed in different output files for different temporal sets, it is easy to compare the listed frequent eventsets. Output files that contain the frequent eventset lists are read separately for two different temporal sets. Files of different temporal sets which have the same file names are compared with each other. As a result, new files are written to output the differences of the files. Figure 5-58 exemplifies the content of the generated output files.

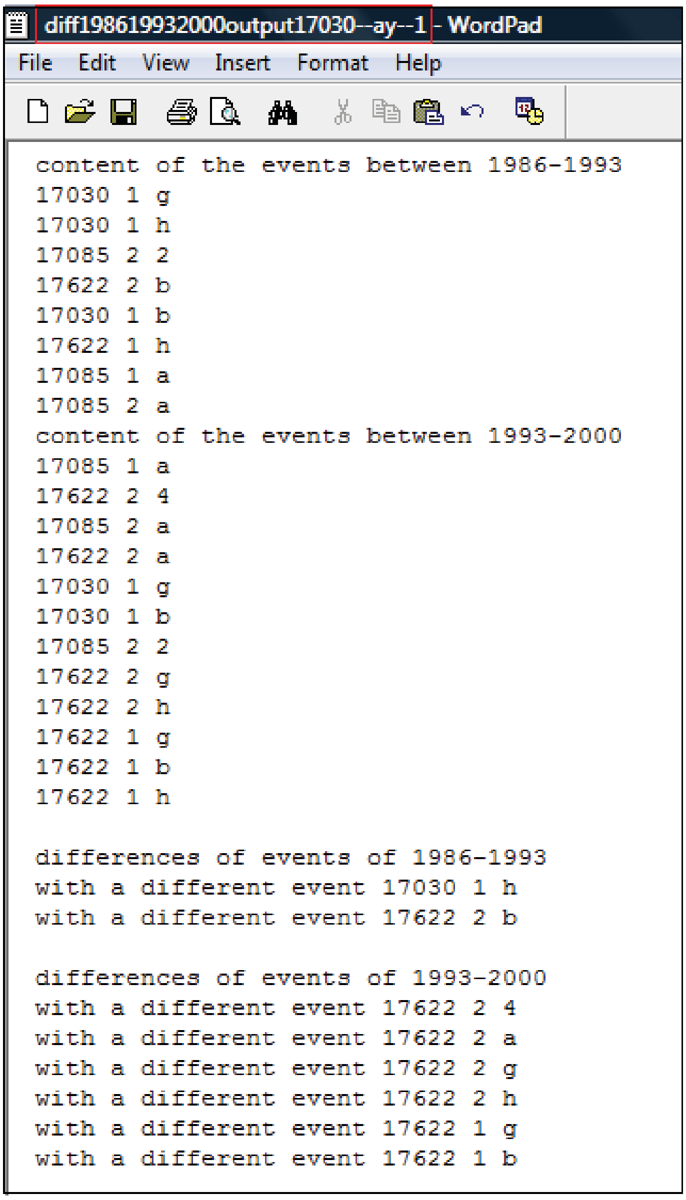


Figure 5-58 Output files which are generated to output the differences

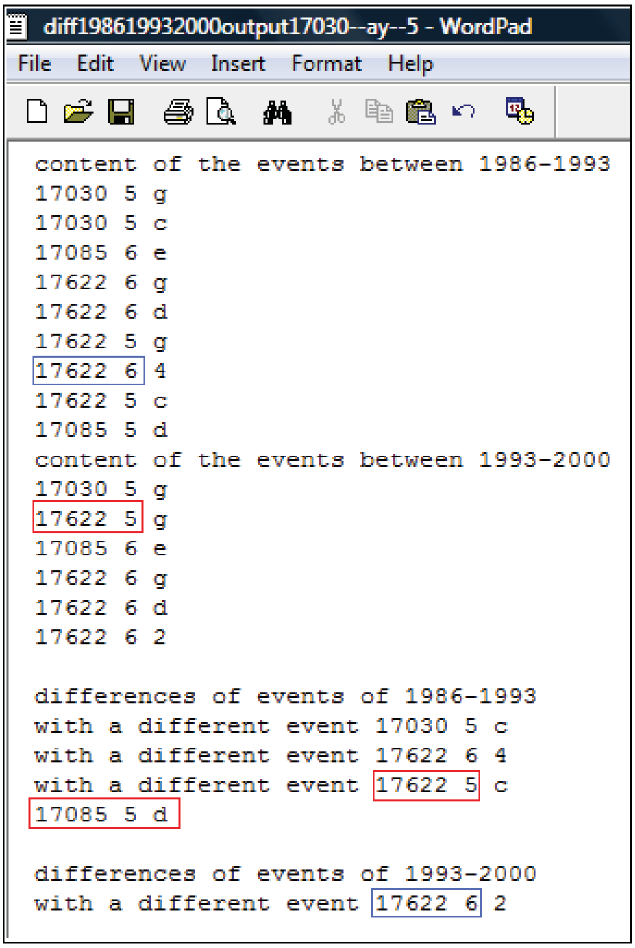
Name of the generated output file is "'diff198619932000output17030—ay--1". Name of the file indicates that temporal sets 1986 to 1993 and 1993 to 2000 are used for comparison. Figured output file is generated from the "output17030—ay--1" files of two different temporal sets. At the beginning of the file frequent eventsets of the output files are listed. Then the differences are documented.

```
differences of events of 1986-1993
with a different event 17030 1 h
with a different event 17622 2 b

differences of events of 1993-2000
with a different event 17622 2 4
with a different event 17622 2 a
with a different event 17622 2 g
with a different event 17622 2 h
with a different event 17622 1 g
with a different event 17622 1 b
```

Figure 5-59 Differences of the eventsets of two different temporal sets

Figure 5-59 shows the section of the output file which documents the differences. Section "differences of events of 1986-19983" contains the eventsets which exist in the output file of temporal set from 1986 to 1993 but do not exist in the output file of temporal set 1986 to 1993. "with a different event" phrase states that there is an eventset in the compared output file for the same weather station for the same month but the event in the eventset is different. "17030 1 h" is stated as a difference of the output file of the temporal set 1986 to 1993. There are "17030 1 g" and "17030 1 b" frequent eventsets in the content of the output file for the temporal set 1993 to 2000. If an eventset is written as a difference without the mentioned phrase, this representation indicates that there is not any frequent eventset in the compared output file with the same station number and month. Following figure exemplifies the explained situation.


diff198619932000output17030--ay--5 - WordPad

File Edit View Insert Format Help

```
content of the events between 1986-1993
17030 5 g
17030 5 c
17085 6 e
17622 6 g
17622 6 d
17622 5 g
17622 6 4
17622 5 c
17085 5 d
content of the events between 1993-2000
17030 5 g
17622 5 g
17085 6 e
17622 6 g
17622 6 d
17622 6 2

differences of events of 1986-1993
with a different event 17030 5 c
with a different event 17622 6 4
with a different event 17622 5 c
17085 5 d

differences of events of 1993-2000
with a different event 17622 6 2
```

Figure 5-60 Difference output file for “17030—ay—5”

As illustrated by Figure 5-60, eventset “17085 5 d” is written as a difference of the temporal set 1986 to 1993. When the content of the output file of the temporal set 1993 to 2000 is analyzed, it is seen that there is not any eventset which starts with “17085 5” (with the same station number and month). However eventset “17622 5 c” is written as a difference with the phrase “with a different event”. Because there is an eventset which starts with “17622 5” in the output file of the temporal set 1993 to 2000.

Next step is the interpretation of the output files. Since there are a lot of output files written to document the differences, only some samples are chosen to denote the affects of the temporal changes on the events. According to the differences documented in the file shown by Figure 5-60, in Bafra (17622) in June average precipitation is between 75 mm and 100 mm (event is 4) for the temporal set 1986 to 1993, on the other hand in Bafra in June average precipitation is between 25 mm and 50 mm (event is 2) for the temporal set 1993 to 2000. There is a significance decrease in the average precipitation. When the precipitation values in two different temporal sets are compared with an un-paired t-test, the two-tailed P values equals to 0.0351. By conventional criteria, this difference is considered to be statistically significant.

```
content of the events between 1986-1993
17033 1 g
17033 1 b
17034 2 b
17624 2 b
17034 1 h
17034 1 b
17624 1 h
17033 1 h
17624 1 b
content of the events between 1993-2000
17033 1 b
17034 1 b
17034 2 5
17624 2 5
17624 1 b
17033 1 g
17034 2 h
17034 2 b
17624 2 g
17624 2 h
17624 2 b
17034 1 g
17624 1 g

differences of events of 1986-1993
with a different event 17034 1 h
with a different event 17624 1 h
with a different event 17033 1 h

differences of events of 1993-2000
with a different event 17034 2 5
with a different event 17624 2 5
with a different event 17034 2 h
with a different event 17624 2 g
with a different event 17624 2 h
with a different event 17034 1 g
with a different event 17624 1 g
```

Figure 5-61 Difference Output file for 17033—ay—1

Figure 5-61 states the snow event change for the stations Ordu (17033), Giresun (17034) and Ünye (17624). Although frequently the number of snowy days is greater than 0 and smaller than 4 in those stations in January for the temporal set 1986 to 1993, number of snowy days is frequently 0 for the same stations in January for the temporal set 1993 to 2000.

```
content of the events between 1986-1993
17033 4 g
17033 4 c
17034 5 g
17624 5 g
17624 5 c
17034 4 g
17034 5 c
17034 4 c
17624 4 g
17624 4 c
content of the events between 1993-2000
17033 4 g
17033 4 c
17034 5 g
17034 5 d
17624 5 g
17034 4 g
17034 5 3
17034 4 c
17624 4 g
17624 4 c
17624 5 d

differences of events of 1986-1993
with a different event 17624 5 c
with a different event 17034 5 c

differences of events of 1993-2000
with a different event 17034 5 d
with a different event 17034 5 3
with a different event 17624 5 d
```

Figure 5-62 Difference Output file for 17033—ay—4

“diff198619932000output17033--ay—4” displayed in Figure 5-62 shows the change of the temperature event by the change of the temporal set. In Ünye (17624) and in Giresun (17034) average temperature is frequently between 10 C° and 15 C° in May for the temporal set 1986 to 1993. But for the next temporal set which is 1993 to 2000, an increase in the average temperature is observed. Frequently average temperature is between 15 C° and 20 C° in May in those stations after 1993.

Another output file which is by Figure 5-63 shows the change of snow event with respect to time. In Sarıkamış (17692) frequently number of snowy days is between 13 and 17 (event l) in April for the temporal set 1993 to 2000. However, number of snowy days is frequently between 4 and 7 (event i) for the temporal set 2000 to 2007 in Sarıkamış in April. There is a decrease in the number of snowy days. When all the number of snowy days of each year in those temporal sets is compared with the unpaired t-test, the two-tailed P value equals to

0.0719. By conventional criteria, this difference is considered to be not quite statistically significant. However, when the frequent events in the temporal sets are considered, comparison results might be different.

```
content of the events between 2000-2007
17656 3 a
17692 3 a
17692 4 i
17692 3 l
17692 4 a
17692 3 m
content of the events between 1993-2000
17656 3 a
17692 3 l
17692 4 l
17692 4 a
17692 3 m
17692 3 a

differences of events of 2000-2007
with a different event 17692 4 i

differences of events of 1993-2000
with a different event 17692 4 l
```

Figure 5-63 Difference output file for station 17656 (Arpacay)

```
content of the events between 2000-2007
17037 6 g
17037 6 3
17088 7 g
17088 7 e
17626 7 e
17088 6 g
17088 7 1
17088 6 d
17626 6 d
content of the events between 1993-2000
17037 6 g
17088 6 g
17088 7 g
17626 7 1
17626 7 e
17088 6 d
17088 7 d

differences of events of 2000-2007
with a different event 17037 6 3
with a different event 17088 7 e
with a different event 17088 7 1
17626 6 d

differences of events of 1993-2000
with a different event 17626 7 1
with a different event 17088 7 d
```

Figure 5-64 Difference output file for station 17037 (Trabzon)

Output file displayed in Figure 5-64 has been written for Trabzon. It is observed an increase in the average temperature in Gümüşhane. Although the average temperature is between 15 C° and 20 C° in July for the temporal set 1993 to 2000, it is between 20 C° and 25 C° for the temporal set 2000 to 2007.

Another output file displayed in Figure 5-65 shows the increase in the average temperature. File is the one written for the station Samsun (17030). In Samsun in July, average temperature is between 20 C° and 25 C° for the temporal set 1993 to 2000. On the other hand, for the temporal set 2000 to 2007, average temperature is over 25 C° in Samsun.

```
content of the events between 2000-2007
17030 6 g
17085 6 e
17085 7 f
17622 7 g
17622 7 1
17085 7 1
17622 7 e
17622 6 g
content of the events between 1993-2000
17030 6 g
17030 6 d
17085 7 1
17085 7 e
17622 7 g
17622 7 e
17085 6 e
17622 7 2
17622 6 g
17622 6 d

differences of events of 2000-2007
with a different event 17085 7 f
with a different event 17622 7 1

differences of events of 1993-2000
with a different event 17030 6 d
with a different event 17085 7 e
with a different event 17622 7 2
with a different event 17622 6 d
```

Figure 5-65 Difference output file for station 17030 (Samsun)

Output file shown in Figure 5-66 also depicts the increase of average temperature for the temporal set 2000 to 2007 when the values are compared with the values of temporal set 1993 to 2000. Average temperature is between less than 5 C° in February in station 17622 (Bafra) after 2000 but it is between 5 C° and 10 C° for the temporal set 1993 to 2000.

```
content of the events between 2000-2007
17085 1 a
17622 2 b
content of the events between 1993-2000
17085 1 a
17622 2 4
17085 2 a
17622 2 a
17030 1 b
17085 2 2
17622 2 g
17622 1 b


differences of events of 2000-2007
with a different event 17622 2 b

differences of events of 1993-2000
with a different event 17622 2 4
17085 2 a
with a different event 17622 2 a
17030 1 b
17085 2 2
17622 1 b
```

Figure 5-66 Difference Output file for 17030—ay—1

Another interesting inspection is about the snow event change in station 17128 (Esenboğa). Output file is displayed by Figure 5-67. Although the number of snowy days is between 10 and 13 (event k) in January in Esenboğa for the temporal set 1986 to 1993, it is only between 0 and 4 for the temporal set 1993 to 2000. There is a significant decrease in the number of snowy days. When the values of two temporal sets are compared with unpaired t-test, the two-tailed P value equals to 0.0453. By conventional criteria, this difference is considered to be statistically significant.

```
content of the events between 1986-1993
17646 1 l
17646 1 m
17128 2 a
17646 1 a
17128 1 k
17128 1 a
content of the events between 1993-2000
17646 1 l
17646 1 m
17128 2 a
17646 1 a
17128 1 2
17128 1 a
17128 1 h

differences of events of 1986-1993
with a different event 17128 1 k

differences of events of 1993-2000
with a different event 17128 1 2
with a different event 17128 1 h
```

Figure 5-67 Snow event change in Esenboğa

It should also be stated that many output files that are written to output the differences of the frequent eventsets are identical with each other for the defined temporal sets. Figure 5-68 is an example of two identical contents of temporal sets. It should be noted that frequencies of the frequent eventsets are not considered during the comparison of the files. Percentage parameter might have been another indicator for the comparison of the output files.

```
content of the events between 1986-1993
17128 3 a
17080 4 g
17135 4 c
17646 4 g
17646 4 b
17080 4 b
17648 4 b
17730 4 b
content of the events between 1993-2000
17128 3 a
17080 4 g
17135 4 c
17646 4 g
17646 4 b
17080 4 b
17648 4 b
17730 4 b

differences of events of 1986-1993

differences of events of 1993-2000
```

Figure 5-68 Identical contents of different temporal sets

CHAPTER 6

CONCLUSION AND FUTURE DIRECTIONS

As already mentioned in the previous sections, spatio-temporal data mining is an emerging research area. Huge collections of spatio-temporal data such as meteorology data contain possibly interesting information and valuable knowledge. By this study, because of the necessity of techniques for the analysis of huge collections of spatio-temporal data, an appoach is proposed for finding interesting patterns. Meterology data that has been collected in various weather stations of Turkey is used to discover spatio-temporal patterns. Although the approach proposed by this study is inspired by J. Wang, et al. [9], it is extended in various aspects. Preprocessing phase is which does not exist in [9] is impelemted. This phase is the first phase of the approach, and it is used for the preparation of the available data set for the implementation. Implementation technique of the proposed approach is completly different than the techniques defined in [9] and [10]. A different candidate generation and mining algorithm is used. Sliding window mechanism that is introduced by this thesis, is one of the most important extensions of the inspired approach. By this sliding window mechanism, the loss of correlations of different temporal intervals is prevented. Applying the approach on a real data set which is really huge, and getting meaningful results were important accomplishments. Defined approach was extended with a new modality named 'Mining Rules'. All the frequent patterns are collected; they are categorized due to their event types, and by a second pass over these categorized frequent patterns the general trend of the events according to the location changes are found out. Experimenting the event changes with respect to temporal changes, demonstrates how the proposed approch can be used for many different and useful investigations.

Future studies on this work can be made in the following aspects:

- A different sample set of stations might be chosen. A different region of the country can be selected. Since each region in Turkey shows various climatic properties, different and more interesting rules can be investigated.
- Sample set might be extended to include the cities which are located at the inner part of the Black Sea Region, in order to observe the significant changes between the coastal and inland parts of the region.
- Some modifications can be introduced to implementation. A better DB may be chosen to handle large datasets. Microsoft Access was not an effective selection. Source code can be parallelized in order to get faster results.

- Classification of the events can be re-arranged with the comments received from Meteorology and Geography experts, to achieve more meaningful results.
- A different technique can be added to the approach to eliminate negligible frequent patterns in the generated results.

REFERENCES

[1] S. U. Calargun, "Fuzzy association rule mining from spatio-temporal data: An analysis of meteorological data in Turkey" Master's thesis, Middle East Technical University, January 2005.

[2] D. J. Peuquet, "A conceptual framework and comparison of spatial data models," Cartographica, vol. 21, no. 4, pp. 66–113, 1984.

[3] J. F. Roddick and M. Spiliopoulou, "A Bibliography of Temporal, Spatial and Spatio-Temporal Data Mining Research", *ACM SIGKDD*, vol. 1, June 1999.

[4] X. Yao, "Research Issues in Spatio-temporal Data Mining", *white paper submitted to the University Consortium for Geographic Information Science (UCGIS) workshop on Geospatial Visualization and Knowledge Discovery, Virginia*. Nov. 18-20, 2003.

[5] K. Koperski and J. Han, "Discovery of spatial association rules in geographic information databases," *4th International Symposium Advances in Spatial Databases*, vol. 951, pp. 47–66, 6–9 1995.

[6] J. Mennis and J. Liu, "Mining Association Rules in Spatio-Temporal Data: An Analysis of Urban Socioeconomic and Land Cover Change", *Transactions in GIS*, vol. 9, issue 1, pages 5-17, January 2005.

[7] L.-Y. Wei and M.-K. Shan, "Mining Temporal Co-orientation Pattern from Spatio-temporal Databases", *PAKDD 2007, LNAI 4426, pp.895-903,* Springer-Verlag Berlin Heidelberg, 2007.

[8] R. B. Subramanyam, A. Goswami and B. Prasad, "Mining fuzzy temporal patterns from process instances with weighted temporal graphs", *International Journal of Data Analysis Techniques and Strategies*, vol. 1, No. 1 / 2008.

[9] J. Wang, W. Hsu, M. Li Lee and J. Wang, "Finding Flow Patterns in Spatio-temporal Databases", *Proceedings of the 16^{th} IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2004)*, 2004.

[10] J. Wang, W. Hsu and M. Li Lee, "Mining Generalized Spatio-Temporal Patterns", *Lecture Notes in Computer Science, LNCS volume 3453/*2005, pp.649-661,Springer Berlin, Heidelberg, 2005.

[11] D. Patel, W. Hsu and M. Li Lee, "Mining Relationships Among Interval-based Events for Classification", *SIGMOD'08, June 9-12, 2008, Vancouver, BC, Canada.*

[12] J. W. Seifert, "Data mining: An overview," tech. rep., Congressional Research Service of the Library of Congress, 2004.

[13] M.-T. Kechadi and M. Bertolotto, "A visual approach for spatio-temporal data mining," IEEE International Conference on Information Reuse and Integration, pp. 504–509, 2006.

[14] H. J. Miller and J. Han, Geographic data mining and knowledge discovery: An overview.Taylor and Francis, 2001.

[15] J. Han and M. Kamber, Data Mining: Concepts and Techniques. Morgan Kaufmann Publisher, Inc, 2001.

[16] “http://www.meteor.gov.tr,” Turkish State Meteorological Service.

[17] S. Shekhar and Y. Huang, Discovery of spatial co-location patterns. *Int. Symposium on Spatial and Temporal Databases*, 2001.

[18] I. Tsoukatos and D. Gunopulos, Efficient mining of spatiotemporal patterns. *Int. Symposium on Spatial and Temporal Databases*, 2001.

[19] J. Wang, W. Hsu and M. Li Lee, “Discovering geographical features for location-based services”, *DASFAA,* 2004.

[20] J. Pei, J. Han, B. Mortazavi, H. Pinto, Q. Chen, U. Dayal, M. and C. Hsu, “PrefixSpan : Mining Sequential Patterns Efficiently by Prefix-Projected Pattern Growth”, 17th International Conference on Data Engineering (ICDE'01), April 02-April 06, ISBN: 0-7695-1001-9

APPENDIX A

OUTPUTS OF THE IMPLEMENTATION

17022-- month —1

```
**************************** Percentage 62.5
month: 1 17070  event is a
month: 2 17070  event is 2
month: 2 17070  event is a
**************************** Percentage 62.5
month: 1 17022  event is b
month: 1 17070  event is a
month: 2 17070  event is 2
**************************** Percentage 62.5
month: 1 17022  event is b
month: 1 17070  event is a
month: 2 17612  event is b
**************************** Percentage 62.5
month: 1 17022  event is b
month: 1 17070  event is a
month: 1 17602  event is b
month: 2 17070  event is a
**************************** Percentage 62.5
month: 1 17022  event is 6
month: 1 17070  event is a
month: 2 17070  event is a
```

17022--1-- month —2

```
**************************** Percentage 68.75
month: 2 17070  event is 2
month: 3 17022  event is g
**************************** Percentage 68.75
month: 2 17070  event is 2
month: 3 17612  event is b
**************************** Percentage 68.75
month: 2 17070  event is 2
month: 2 17070  event is a
month: 3 17679  event is b
**************************** Percentage 68.75
month: 2 17070  event is 2
month: 2 17070  event is a
month: 3 17680  event is g
**************************** Percentage 68.75
month: 2 17070  event is 2
month: 2 17070  event is a
month: 3 17680  event is b
```

17022-- month —2

```
**************************** Percentage 62.5
month: 2 17070  event is 2
month: 3 17602  event is b
**************************** Percentage 62.5
month: 2 17070  event is 2
```

```
month: 3 17612  event is b
**************************** Percentage 62.5
month: 2 17070  event is a
month: 3 17602  event is b
**************************** Percentage 62.5
month: 2 17070  event is a
month: 3 17612  event is b
```

17022--2-- month —3

```
**************************** Percentage 75.0
month: 3 17602  event is b
month: 3 17612  event is b
month: 4 17022  event is g
**************************** Percentage 68.75
month: 3 17602  event is b
month: 3 17612  event is b
month: 4 17679  event is c
**************************** Percentage 75.0
month: 3 17602  event is b
month: 3 17612  event is b
month: 4 17680  event is g
**************************** Percentage 75.0
month: 3 17602  event is b
month: 3 17612  event is b
month: 4 17070  event is g
```

17022-- month —3

```
**************************** Percentage 68.75
month: 3 17022  event is g
month: 3 17602  event is b
month: 3 17612  event is b
month: 4 17070  event is g
**************************** Percentage 62.5
month: 3 17022  event is b
month: 3 17602  event is b
month: 3 17612  event is b
month: 4 17070  event is g
```

17022--3-- month —4

```
**************************** Percentage 75.0
month: 4 17070  event is g
month: 5 17679  event is d
**************************** Percentage 81.25
month: 4 17070  event is g
month: 5 17680  event is d
**************************** Percentage 68.75
month: 4 17070  event is g
month: 4 17602  event is c
month: 5 17022  event is g
**************************** Percentage 68.75
month: 4 17070  event is g
month: 4 17602  event is c
month: 5 17680  event is g
**************************** Percentage 68.75
month: 4 17070  event is g
```

```
month: 4 17602  event is c
month: 5 17070  event is g
**************************** Percentage 68.75
month: 4 17070  event is g
month: 4 17070  event is b
month: 5 17022  event is g
**************************** Percentage 68.75
month: 4 17070  event is g
month: 4 17070  event is b
month: 5 17070  event is c
```

17022-- month —4

```
**************************** Percentage 62.5
month: 4 17022  event is g
month: 4 17070  event is g
month: 5 17612  event is c
**************************** Percentage 75.0
month: 4 17022  event is g
month: 4 17022  event is c
month: 4 17070  event is g
month: 5 17070  event is g
**************************** Percentage 68.75
month: 4 17022  event is g
month: 4 17022  event is c
month: 4 17070  event is g
month: 5 17070  event is c
**************************** Percentage 62.5
month: 4 17022  event is g
month: 4 17022  event is c
month: 4 17602  event is c
month: 5 17070  event is g
**************************** Percentage 62.5
month: 4 17022  event is c
month: 4 17070  event is g
month: 4 17602  event is c
month: 5 17070  event is g
**************************** Percentage 62.5
month: 4 17022  event is g
month: 4 17070  event is g
month: 4 17070  event is b
month: 5 17070  event is c
```

17022--4-- month —5

```
**************************** Percentage 87.5
month: 5 17070  event is g
month: 5 17070  event is c
month: 6 17022  event is g
**************************** Percentage 87.5
month: 5 17070  event is g
month: 5 17070  event is c
month: 6 17612  event is d
**************************** Percentage 81.25
month: 5 17070  event is g
month: 5 17070  event is c
month: 6 17680  event is e
**************************** Percentage 87.5
month: 5 17070  event is g
```

```
month: 5 17070  event is c
month: 6 17070  event is g
**************************** Percentage 87.5
month: 5 17070  event is g
month: 5 17070  event is c
month: 6 17070  event is d
**************************** Percentage 68.75
month: 5 17070  event is g
month: 6 17022  event is d
**************************** Percentage 68.75
month: 5 17070  event is g
month: 6 17679  event is l
```

17022-- month —5

```
**************************** Percentage 87.5
month: 5 17022  event is g
month: 5 17070  event is g
month: 5 17070  event is c
month: 6 17070  event is g
**************************** Percentage 87.5
month: 5 17022  event is g
month: 5 17070  event is g
month: 5 17070  event is c
month: 6 17612  event is d
**************************** Percentage 62.5
month: 5 17022  event is g
month: 5 17070  event is g
month: 5 17612  event is c
month: 6 17070  event is d
**************************** Percentage 62.5
month: 5 17022  event is g
month: 5 17070  event is g
month: 5 17612  event is c
month: 6 17612  event is d
**************************** Percentage 62.5
month: 5 17022  event is g
month: 5 17070  event is c
month: 5 17612  event is c
month: 6 17070  event is d
**************************** Percentage 62.5
month: 5 17022  event is g
month: 5 17070  event is c
month: 5 17612  event is c
month: 6 17612  event is d
**************************** Percentage 62.5
month: 5 17070  event is g
month: 5 17070  event is c
month: 5 17612  event is c
month: 6 17070  event is d
**************************** Percentage 75.0
month: 5 17022  event is g
month: 5 17070  event is g
month: 5 17070  event is c
month: 6 17602  event is d
```

17022--5-- month —6

```
**************************** Percentage 93.75
```

```
month: 6 17070  event is g
month: 6 17070  event is d
month: 6 17612  event is d
month: 7 17022  event is g
**************************** Percentage 87.5
month: 6 17070  event is g
month: 6 17070  event is d
month: 6 17612  event is d
month: 7 17022  event is e
**************************** Percentage 81.25
month: 6 17070  event is g
month: 6 17070  event is d
month: 6 17612  event is d
month: 7 17612  event is e
**************************** Percentage 93.75
month: 6 17070  event is g
month: 6 17070  event is d
month: 6 17612  event is d
month: 7 17679  event is 1
**************************** Percentage 87.5
month: 6 17070  event is g
month: 6 17070  event is d
month: 6 17612  event is d
month: 7 17679  event is e
**************************** Percentage 81.25
month: 6 17070  event is g
month: 6 17070  event is d
month: 6 17612  event is d
month: 7 17680  event is 1
**************************** Percentage 93.75
month: 6 17070  event is g
month: 6 17070  event is d
month: 6 17612  event is d
month: 7 17070  event is g
**************************** Percentage 75.0
month: 6 17070  event is g
month: 6 17070  event is d
month: 6 17602  event is d
month: 7 17022  event is e
**************************** Percentage 68.75
month: 6 17070  event is g
month: 6 17070  event is d
month: 6 17602  event is d
month: 7 17612  event is e
**************************** Percentage 81.25
month: 6 17070  event is g
month: 6 17070  event is d
month: 6 17602  event is d
month: 7 17679  event is 1
**************************** Percentage 75.0
month: 6 17070  event is g
month: 6 17070  event is d
month: 6 17602  event is d
month: 7 17679  event is e
**************************** Percentage 68.75
month: 6 17070  event is g
month: 6 17070  event is d
month: 6 17602  event is d
month: 7 17680  event is 1
**************************** Percentage 68.75
```

```
month: 6 17070  event is g
month: 6 17070  event is d
month: 6 17612  event is d
month: 7 17070  event is d
**************************** Percentage 81.25
month: 6 17070  event is g
month: 6 17602  event is d
month: 6 17612  event is d
month: 7 17022  event is g
**************************** Percentage 75.0
month: 6 17070  event is g
month: 6 17602  event is d
month: 6 17612  event is d
month: 7 17022  event is e
**************************** Percentage 81.25
month: 6 17070  event is g
month: 6 17602  event is d
month: 6 17612  event is d
month: 7 17679  event is 1
**************************** Percentage 75.0
month: 6 17070  event is g
month: 6 17602  event is d
month: 6 17612  event is d
month: 7 17679  event is e
**************************** Percentage 68.75
month: 6 17070  event is g
month: 6 17602  event is d
month: 6 17612  event is d
month: 7 17680  event is 1
**************************** Percentage 81.25
month: 6 17070  event is d
month: 6 17602  event is d
month: 6 17612  event is d
month: 7 17022  event is g
**************************** Percentage 75.0
month: 6 17070  event is d
month: 6 17602  event is d
month: 6 17612  event is d
month: 7 17022  event is e
**************************** Percentage 81.25
month: 6 17070  event is d
month: 6 17602  event is d
month: 6 17612  event is d
month: 7 17679  event is 1
**************************** Percentage 75.0
month: 6 17070  event is d
month: 6 17602  event is d
month: 6 17612  event is d
month: 7 17679  event is e
**************************** Percentage 81.25
month: 6 17070  event is d
month: 6 17602  event is d
month: 6 17612  event is d
month: 7 17680  event is g
**************************** Percentage 68.75
month: 6 17070  event is d
month: 6 17602  event is d
month: 6 17612  event is d
month: 7 17680  event is 1
**************************** Percentage 81.25
```

```
month: 6 17070  event is d
month: 6 17602  event is d
month: 6 17612  event is d
month: 7 17070  event is g
```

17022-- month —6

```
**************************** Percentage 93.75
month: 6 17022  event is g
month: 6 17070  event is g
month: 6 17070  event is d
month: 7 17070  event is g
**************************** Percentage 93.75
month: 6 17022  event is g
month: 6 17070  event is g
month: 6 17070  event is d
month: 7 17602  event is e
**************************** Percentage 81.25
month: 6 17022  event is g
month: 6 17070  event is g
month: 6 17070  event is d
month: 7 17612  event is e
**************************** Percentage 93.75
month: 6 17022  event is g
month: 6 17070  event is d
month: 6 17612  event is d
month: 7 17070  event is g
**************************** Percentage 93.75
month: 6 17022  event is g
month: 6 17070  event is d
month: 6 17612  event is d
month: 7 17602  event is e
**************************** Percentage 81.25
month: 6 17022  event is g
month: 6 17070  event is d
month: 6 17612  event is d
month: 7 17612  event is e
**************************** Percentage 93.75
month: 6 17070  event is g
month: 6 17070  event is d
month: 6 17612  event is d
month: 7 17070  event is g
**************************** Percentage 93.75
month: 6 17070  event is g
month: 6 17070  event is d
month: 6 17612  event is d
month: 7 17602  event is e
**************************** Percentage 81.25
month: 6 17070  event is g
month: 6 17070  event is d
month: 6 17612  event is d
month: 7 17612  event is e
**************************** Percentage 68.75
month: 6 17022  event is g
month: 6 17022  event is d
month: 6 17070  event is g
month: 7 17602  event is e
**************************** Percentage 68.75
month: 6 17022  event is g
```

```
month: 6 17022  event is d
month: 6 17070  event is d
month: 7 17602  event is e
**************************** Percentage 68.75
month: 6 17022  event is d
month: 6 17070  event is g
month: 6 17070  event is d
month: 7 17070  event is g
**************************** Percentage 68.75
month: 6 17022  event is d
month: 6 17070  event is g
month: 6 17070  event is d
month: 7 17602  event is e
**************************** Percentage 68.75
month: 6 17022  event is g
month: 6 17022  event is d
month: 6 17602  event is d
month: 7 17602  event is e
**************************** Percentage 81.25
month: 6 17022  event is g
month: 6 17070  event is g
month: 6 17602  event is d
month: 7 17070  event is g
**************************** Percentage 81.25
month: 6 17022  event is g
month: 6 17070  event is g
month: 6 17602  event is d
month: 7 17602  event is e
**************************** Percentage 68.75
month: 6 17022  event is d
month: 6 17070  event is g
month: 6 17602  event is d
month: 7 17070  event is g
**************************** Percentage 68.75
month: 6 17022  event is d
month: 6 17070  event is g
month: 6 17602  event is d
month: 7 17602  event is e
**************************** Percentage 81.25
month: 6 17022  event is g
month: 6 17070  event is d
month: 6 17602  event is d
month: 7 17070  event is g
**************************** Percentage 68.75
month: 6 17022  event is g
month: 6 17070  event is d
month: 6 17602  event is d
month: 7 17612  event is e
**************************** Percentage 68.75
month: 6 17022  event is d
month: 6 17070  event is d
month: 6 17602  event is d
month: 7 17070  event is g
**************************** Percentage 68.75
month: 6 17022  event is d
month: 6 17070  event is d
month: 6 17602  event is d
month: 7 17602  event is e
**************************** Percentage 62.5
month: 6 17070  event is g
```

```
month: 6 17070  event is d
month: 6 17602  event is d
month: 7 17070  event is d
**************************** Percentage 81.25
month: 6 17070  event is g
month: 6 17070  event is d
month: 6 17602  event is d
month: 7 17602  event is e
**************************** Percentage 68.75
month: 6 17070  event is g
month: 6 17070  event is d
month: 6 17602  event is d
month: 7 17612  event is e
**************************** Percentage 68.75
month: 6 17022  event is d
month: 6 17070  event is g
month: 6 17612  event is d
month: 7 17602  event is e
**************************** Percentage 68.75
month: 6 17022  event is d
month: 6 17070  event is d
month: 6 17612  event is d
month: 7 17070  event is g
**************************** Percentage 68.75
month: 6 17022  event is d
month: 6 17070  event is d
month: 6 17612  event is d
month: 7 17602  event is e
**************************** Percentage 68.75
month: 6 17070  event is g
month: 6 17070  event is d
month: 6 17612  event is d
month: 7 17070  event is d
**************************** Percentage 81.25
month: 6 17022  event is g
month: 6 17602  event is d
month: 6 17612  event is d
month: 7 17070  event is g
**************************** Percentage 81.25
month: 6 17022  event is g
month: 6 17602  event is d
month: 6 17612  event is d
month: 7 17602  event is e
**************************** Percentage 68.75
month: 6 17022  event is g
month: 6 17602  event is d
month: 6 17612  event is d
month: 7 17612  event is e
**************************** Percentage 68.75
month: 6 17022  event is d
month: 6 17602  event is d
month: 6 17612  event is d
month: 7 17070  event is g
**************************** Percentage 68.75
month: 6 17022  event is d
month: 6 17602  event is d
month: 6 17612  event is d
month: 7 17602  event is e
**************************** Percentage 62.5
month: 6 17070  event is g
```

```
month: 6 17602  event is d
month: 6 17612  event is d
month: 7 17070  event is d
**************************** Percentage 81.25
month: 6 17070  event is g
month: 6 17602  event is d
month: 6 17612  event is d
month: 7 17602  event is e
**************************** Percentage 68.75
month: 6 17070  event is g
month: 6 17602  event is d
month: 6 17612  event is d
month: 7 17612  event is e
**************************** Percentage 81.25
month: 6 17070  event is d
month: 6 17602  event is d
month: 6 17612  event is d
month: 7 17070  event is g
**************************** Percentage 62.5
month: 6 17070  event is d
month: 6 17602  event is d
month: 6 17612  event is d
month: 7 17070  event is d
**************************** Percentage 81.25
month: 6 17070  event is d
month: 6 17602  event is d
month: 6 17612  event is d
month: 7 17602  event is e
```

17022--6-- month —7

```
**************************** Percentage 75.0
month: 7 17070  event is g
month: 7 17602  event is e
month: 8 17022  event is e
**************************** Percentage 68.75
month: 7 17070  event is g
month: 7 17602  event is e
month: 8 17680  event is e
**************************** Percentage 81.25
month: 7 17070  event is g
month: 7 17602  event is e
month: 7 17612  event is e
month: 8 17022  event is g
**************************** Percentage 75.0
month: 7 17070  event is g
month: 7 17602  event is e
month: 7 17612  event is e
month: 8 17679  event is 1
**************************** Percentage 75.0
month: 7 17070  event is g
month: 7 17602  event is e
month: 7 17612  event is e
month: 8 17679  event is e
**************************** Percentage 81.25
month: 7 17070  event is g
month: 7 17602  event is e
month: 7 17612  event is e
month: 8 17680  event is g
```

```
**************************** Percentage 68.75
month: 7 17070  event is g
month: 7 17602  event is e
month: 7 17612  event is e
month: 8 17680  event is 1
**************************** Percentage 81.25
month: 7 17070  event is g
month: 7 17602  event is e
month: 7 17612  event is e
month: 8 17070  event is g
**************************** Percentage 75.0
month: 7 17070  event is g
month: 7 17602  event is e
month: 7 17612  event is e
month: 8 17070  event is d
**************************** Percentage 68.75
month: 7 17070  event is g
month: 7 17070  event is d
month: 7 17602  event is e
month: 8 17680  event is g
**************************** Percentage 68.75
month: 7 17070  event is g
month: 7 17070  event is d
month: 7 17602  event is e
month: 8 17070  event is g
**************************** Percentage 68.75
month: 7 17070  event is g
month: 7 17602  event is e
month: 8 17070  event is 1
```

17022-- month —7

```
**************************** Percentage 75.0
month: 7 17022  event is g
month: 7 17022  event is e
month: 7 17070  event is g
month: 8 17070  event is d
**************************** Percentage 87.5
month: 7 17022  event is g
month: 7 17022  event is e
month: 7 17602  event is e
month: 8 17070  event is g
**************************** Percentage 75.0
month: 7 17022  event is g
month: 7 17022  event is e
month: 7 17602  event is e
month: 8 17070  event is d
**************************** Percentage 81.25
month: 7 17022  event is g
month: 7 17022  event is e
month: 7 17602  event is e
month: 8 17602  event is e
**************************** Percentage 93.75
month: 7 17022  event is g
month: 7 17070  event is g
month: 7 17602  event is e
month: 8 17070  event is g
**************************** Percentage 75.0
month: 7 17022  event is g
month: 7 17070  event is g
month: 7 17602  event is e
```

```
month: 8 17070  event is d
**************************** Percentage 87.5
month: 7 17022  event is g
month: 7 17070  event is g
month: 7 17602  event is e
month: 8 17602  event is e
**************************** Percentage 87.5
month: 7 17022  event is e
month: 7 17070  event is g
month: 7 17602  event is e
month: 8 17070  event is g
**************************** Percentage 75.0
month: 7 17022  event is e
month: 7 17070  event is g
month: 7 17602  event is e
month: 8 17070  event is d
**************************** Percentage 81.25
month: 7 17022  event is e
month: 7 17070  event is g
month: 7 17602  event is e
month: 8 17602  event is e
**************************** Percentage 68.75
month: 7 17022  event is g
month: 7 17022  event is e
month: 7 17612  event is e
month: 8 17070  event is d
**************************** Percentage 75.0
month: 7 17022  event is g
month: 7 17022  event is e
month: 7 17612  event is e
month: 8 17602  event is e
**************************** Percentage 68.75
month: 7 17022  event is g
month: 7 17070  event is g
month: 7 17612  event is e
month: 8 17070  event is d
**************************** Percentage 81.25
month: 7 17022  event is e
month: 7 17070  event is g
month: 7 17612  event is e
month: 8 17070  event is g
**************************** Percentage 68.75
month: 7 17022  event is e
month: 7 17070  event is g
month: 7 17612  event is e
month: 8 17070  event is d
**************************** Percentage 81.25
month: 7 17022  event is g
month: 7 17602  event is e
month: 7 17612  event is e
month: 8 17070  event is g
**************************** Percentage 68.75
month: 7 17022  event is g
month: 7 17602  event is e
month: 7 17612  event is e
month: 8 17070  event is d
**************************** Percentage 81.25
month: 7 17022  event is e
month: 7 17602  event is e
month: 7 17612  event is e
```

```
month: 8 17070  event is g
**************************** Percentage 68.75
month: 7 17022  event is e
month: 7 17602  event is e
month: 7 17612  event is e
month: 8 17070  event is d
**************************** Percentage 75.0
month: 7 17022  event is e
month: 7 17602  event is e
month: 7 17612  event is e
month: 8 17602  event is e
**************************** Percentage 81.25
month: 7 17070  event is g
month: 7 17602  event is e
month: 7 17612  event is e
month: 8 17070  event is g
**************************** Percentage 68.75
month: 7 17070  event is g
month: 7 17602  event is e
month: 7 17612  event is e
month: 8 17070  event is d
**************************** Percentage 75.0
month: 7 17070  event is g
month: 7 17602  event is e
month: 7 17612  event is e
month: 8 17602  event is e
**************************** Percentage 62.5
month: 7 17022  event is e
month: 7 17070  event is g
month: 7 17070  event is d
month: 8 17070  event is g
**************************** Percentage 68.75
month: 7 17022  event is g
month: 7 17070  event is d
month: 7 17602  event is e
month: 8 17070  event is g
**************************** Percentage 62.5
month: 7 17022  event is e
month: 7 17070  event is d
month: 7 17602  event is e
month: 8 17070  event is g
**************************** Percentage 62.5
month: 7 17022  event is g
month: 7 17070  event is g
month: 7 17070  event is d
month: 8 17602  event is e
**************************** Percentage 62.5
month: 7 17022  event is g
month: 7 17070  event is g
month: 7 17602  event is e
month: 8 17070  event is 1
**************************** Percentage 62.5
month: 7 17022  event is g
month: 7 17070  event is d
month: 7 17602  event is e
month: 8 17602  event is e
**************************** Percentage 62.5
month: 7 17070  event is g
month: 7 17070  event is d
month: 7 17602  event is e
```

```
month: 8 17602  event is e
```

17022--7-- month —8

```
**************************** Percentage 68.75
month: 8 17070  event is g
month: 8 17070  event is d
month: 9 17679  event is 1
**************************** Percentage 75.0
month: 8 17070  event is g
month: 8 17602  event is e
month: 9 17679  event is 1
**************************** Percentage 75.0
month: 8 17070  event is g
month: 8 17070  event is d
month: 8 17602  event is e
month: 9 17612  event is d
**************************** Percentage 75.0
month: 8 17070  event is g
month: 8 17070  event is d
month: 8 17602  event is e
month: 9 17680  event is g
**************************** Percentage 68.75
month: 8 17070  event is g
month: 8 17070  event is d
month: 8 17602  event is e
month: 9 17680  event is 1
**************************** Percentage 75.0
month: 8 17070  event is g
month: 8 17070  event is d
month: 8 17602  event is e
month: 9 17070  event is g
**************************** Percentage 68.75
month: 8 17070  event is g
month: 8 17070  event is d
month: 8 17602  event is e
month: 9 17070  event is d
**************************** Percentage 68.75
month: 8 17070  event is g
month: 8 17070  event is 1
month: 8 17602  event is e
month: 9 17022  event is d
**************************** Percentage 68.75
month: 8 17070  event is g
month: 8 17070  event is 1
month: 8 17602  event is e
month: 9 17680  event is g
```

17022-- month —8

```
**************************** Percentage 62.5
month: 8 17022  event is g
month: 8 17070  event is g
month: 9 17070  event is 1
**************************** Percentage 75.0
month: 8 17022  event is g
month: 8 17022  event is e
month: 8 17070  event is g
month: 9 17070  event is g
```

```
**************************** Percentage 62.5
month: 8 17022  event is g
month: 8 17022  event is e
month: 8 17070  event is g
month: 9 17070  event is d
**************************** Percentage 75.0
month: 8 17022  event is g
month: 8 17022  event is e
month: 8 17070  event is g
month: 9 17612  event is d
**************************** Percentage 75.0
month: 8 17022  event is g
month: 8 17070  event is g
month: 8 17070  event is d
month: 9 17612  event is d
**************************** Percentage 62.5
month: 8 17022  event is g
month: 8 17022  event is e
month: 8 17602  event is e
month: 9 17070  event is d
**************************** Percentage 75.0
month: 8 17022  event is g
month: 8 17022  event is e
month: 8 17602  event is e
month: 9 17612  event is d
**************************** Percentage 75.0
month: 8 17022  event is g
month: 8 17070  event is g
month: 8 17602  event is e
month: 9 17602  event is d
**************************** Percentage 87.5
month: 8 17022  event is g
month: 8 17070  event is g
month: 8 17602  event is e
month: 9 17612  event is d
**************************** Percentage 62.5
month: 8 17022  event is e
month: 8 17070  event is g
month: 8 17602  event is e
month: 9 17070  event is d
**************************** Percentage 75.0
month: 8 17022  event is e
month: 8 17070  event is g
month: 8 17602  event is e
month: 9 17612  event is d
**************************** Percentage 68.75
month: 8 17022  event is g
month: 8 17070  event is d
month: 8 17602  event is e
month: 9 17070  event is g
**************************** Percentage 62.5
month: 8 17022  event is g
month: 8 17070  event is d
month: 8 17602  event is e
month: 9 17070  event is d
**************************** Percentage 68.75
month: 8 17022  event is g
month: 8 17070  event is d
month: 8 17602  event is e
month: 9 17612  event is d
```

```
**************************** Percentage 68.75
month: 8 17070  event is g
month: 8 17070  event is d
month: 8 17602  event is e
month: 9 17612  event is d
**************************** Percentage 62.5
month: 8 17022  event is g
month: 8 17022  event is e
month: 8 17612  event is e
month: 9 17612  event is d
**************************** Percentage 62.5
month: 8 17022  event is g
month: 8 17070  event is g
month: 8 17612  event is e
month: 9 17070  event is g
**************************** Percentage 62.5
month: 8 17022  event is g
month: 8 17070  event is g
month: 8 17612  event is e
month: 9 17612  event is d
**************************** Percentage 62.5
month: 8 17022  event is e
month: 8 17070  event is g
month: 8 17612  event is e
month: 9 17612  event is d
**************************** Percentage 62.5
month: 8 17022  event is g
month: 8 17602  event is e
month: 8 17612  event is e
month: 9 17070  event is g
**************************** Percentage 62.5
month: 8 17022  event is g
month: 8 17602  event is e
month: 8 17612  event is e
month: 9 17612  event is d
**************************** Percentage 62.5
month: 8 17022  event is e
month: 8 17602  event is e
month: 8 17612  event is e
month: 9 17070  event is g
**************************** Percentage 62.5
month: 8 17022  event is e
month: 8 17602  event is e
month: 8 17612  event is e
month: 9 17612  event is d
**************************** Percentage 62.5
month: 8 17070  event is g
month: 8 17602  event is e
month: 8 17612  event is e
month: 9 17070  event is g
**************************** Percentage 62.5
month: 8 17070  event is g
month: 8 17602  event is e
month: 8 17612  event is e
month: 9 17612  event is d
**************************** Percentage 62.5
month: 8 17022  event is g
month: 8 17070  event is g
month: 8 17070  event is 1
month: 9 17602  event is d
```

```
**************************** Percentage 62.5
month: 8 17022  event is g
month: 8 17070  event is g
month: 8 17070  event is 1
month: 9 17612  event is d
**************************** Percentage 62.5
month: 8 17022  event is g
month: 8 17070  event is 1
month: 8 17602  event is e
month: 9 17070  event is g
**************************** Percentage 62.5
month: 8 17022  event is g
month: 8 17070  event is 1
month: 8 17602  event is e
month: 9 17602  event is d
**************************** Percentage 62.5
month: 8 17022  event is g
month: 8 17070  event is 1
month: 8 17602  event is e
month: 9 17612  event is d
**************************** Percentage 62.5
month: 8 17070  event is g
month: 8 17070  event is 1
month: 8 17602  event is e
month: 9 17070  event is g
**************************** Percentage 62.5
month: 8 17070  event is g
month: 8 17070  event is 1
month: 8 17602  event is e
month: 9 17602  event is d
**************************** Percentage 62.5
month: 8 17070  event is g
month: 8 17070  event is 1
month: 8 17602  event is e
month: 9 17612  event is d
```

17022--8-- month —9

```
**************************** Percentage 75.0
month: 9 17070  event is g
month: 9 17612  event is d
month: 10 17679  event is 1
**************************** Percentage 68.75
month: 9 17070  event is g
month: 9 17612  event is d
month: 10 17680  event is c
**************************** Percentage 75.0
month: 9 17070  event is g
month: 9 17070  event is d
month: 9 17612  event is d
month: 10 17022  event is g
**************************** Percentage 68.75
month: 9 17070  event is g
month: 9 17070  event is d
month: 9 17612  event is d
month: 10 17612  event is c
**************************** Percentage 75.0
month: 9 17070  event is g
month: 9 17070  event is d
```

```
month: 9 17612  event is d
month: 10 17680  event is g
**************************** Percentage 75.0
month: 9 17070  event is g
month: 9 17070  event is d
month: 9 17612  event is d
month: 10 17070  event is g
**************************** Percentage 81.25
month: 9 17070  event is g
month: 9 17602  event is d
month: 9 17612  event is d
month: 10 17022  event is g
**************************** Percentage 75.0
month: 9 17070  event is g
month: 9 17602  event is d
month: 9 17612  event is d
month: 10 17612  event is c
**************************** Percentage 68.75
month: 9 17070  event is g
month: 9 17602  event is d
month: 9 17612  event is d
month: 10 17679  event is c
**************************** Percentage 81.25
month: 9 17070  event is g
month: 9 17602  event is d
month: 9 17612  event is d
month: 10 17680  event is g
**************************** Percentage 81.25
month: 9 17070  event is g
month: 9 17602  event is d
month: 9 17612  event is d
month: 10 17070  event is g
```

17022-- month —9

```
**************************** Percentage 93.75
month: 9 17022  event is g
month: 9 17022  event is d
month: 9 17070  event is g
month: 10 17070  event is g
**************************** Percentage 87.5
month: 9 17022  event is g
month: 9 17022  event is d
month: 9 17070  event is g
month: 10 17612  event is c
**************************** Percentage 62.5
month: 9 17022  event is g
month: 9 17022  event is d
month: 9 17070  event is 1
month: 10 17612  event is c
**************************** Percentage 62.5
month: 9 17022  event is g
month: 9 17070  event is g
month: 9 17070  event is 1
month: 10 17612  event is c
**************************** Percentage 62.5
month: 9 17022  event is d
month: 9 17070  event is g
month: 9 17070  event is 1
```

```
month: 10 17070  event is g
**************************** Percentage 62.5
month: 9 17022  event is d
month: 9 17070  event is g
month: 9 17070  event is 1
month: 10 17612  event is c
**************************** Percentage 75.0
month: 9 17022  event is g
month: 9 17022  event is d
month: 9 17070  event is d
month: 10 17070  event is g
**************************** Percentage 68.75
month: 9 17022  event is g
month: 9 17022  event is d
month: 9 17070  event is d
month: 10 17612  event is c
**************************** Percentage 68.75
month: 9 17022  event is d
month: 9 17070  event is g
month: 9 17070  event is d
month: 10 17612  event is c
**************************** Percentage 81.25
month: 9 17022  event is g
month: 9 17022  event is d
month: 9 17602  event is d
month: 10 17070  event is g
**************************** Percentage 75.0
month: 9 17022  event is g
month: 9 17022  event is d
month: 9 17602  event is d
month: 10 17612  event is c
**************************** Percentage 81.25
month: 9 17022  event is d
month: 9 17070  event is g
month: 9 17602  event is d
month: 10 17070  event is g
**************************** Percentage 75.0
month: 9 17022  event is d
month: 9 17070  event is g
month: 9 17602  event is d
month: 10 17612  event is c
**************************** Percentage 93.75
month: 9 17022  event is g
month: 9 17022  event is d
month: 9 17612  event is d
month: 10 17070  event is g
**************************** Percentage 68.75
month: 9 17022  event is g
month: 9 17022  event is d
month: 9 17612  event is d
month: 10 17070  event is c
**************************** Percentage 87.5
month: 9 17022  event is g
month: 9 17022  event is d
month: 9 17612  event is d
month: 10 17612  event is c
**************************** Percentage 87.5
month: 9 17022  event is g
month: 9 17070  event is g
month: 9 17612  event is d
```

```
month: 10 17612  event is c
**************************** Percentage 93.75
month: 9 17022  event is d
month: 9 17070  event is g
month: 9 17612  event is d
month: 10 17070  event is g
**************************** Percentage 68.75
month: 9 17022  event is d
month: 9 17070  event is g
month: 9 17612  event is d
month: 10 17070  event is c
**************************** Percentage 87.5
month: 9 17022  event is d
month: 9 17070  event is g
month: 9 17612  event is d
month: 10 17612  event is c
**************************** Percentage 62.5
month: 9 17022  event is g
month: 9 17070  event is 1
month: 9 17612  event is d
month: 10 17612  event is c
**************************** Percentage 62.5
month: 9 17022  event is d
month: 9 17070  event is 1
month: 9 17612  event is d
month: 10 17070  event is g
**************************** Percentage 62.5
month: 9 17022  event is d
month: 9 17070  event is 1
month: 9 17612  event is d
month: 10 17612  event is c
**************************** Percentage 62.5
month: 9 17070  event is g
month: 9 17070  event is 1
month: 9 17612  event is d
month: 10 17070  event is g
**************************** Percentage 62.5
month: 9 17070  event is g
month: 9 17070  event is 1
month: 9 17612  event is d
month: 10 17612  event is c
**************************** Percentage 75.0
month: 9 17022  event is g
month: 9 17070  event is d
month: 9 17612  event is d
month: 10 17070  event is g
**************************** Percentage 68.75
month: 9 17022  event is d
month: 9 17070  event is d
month: 9 17612  event is d
month: 10 17612  event is c
**************************** Percentage 68.75
month: 9 17070  event is g
month: 9 17070  event is d
month: 9 17612  event is d
month: 10 17612  event is c
**************************** Percentage 81.25
month: 9 17022  event is g
month: 9 17602  event is d
month: 9 17612  event is d
```

```
month: 10 17070  event is g
**************************** Percentage 75.0
month: 9 17022  event is g
month: 9 17602  event is d
month: 9 17612  event is d
month: 10 17612  event is c
**************************** Percentage 81.25
month: 9 17022  event is d
month: 9 17602  event is d
month: 9 17612  event is d
month: 10 17070  event is g
**************************** Percentage 75.0
month: 9 17022  event is d
month: 9 17602  event is d
month: 9 17612  event is d
month: 10 17612  event is c
**************************** Percentage 81.25
month: 9 17070  event is g
month: 9 17602  event is d
month: 9 17612  event is d
month: 10 17070  event is g
**************************** Percentage 75.0
month: 9 17070  event is g
month: 9 17602  event is d
month: 9 17612  event is d
month: 10 17612  event is c
**************************** Percentage 62.5
month: 9 17022  event is g
month: 9 17022  event is d
month: 9 17070  event is g
month: 10 17602  event is c
**************************** Percentage 62.5
month: 9 17022  event is g
month: 9 17022  event is d
month: 9 17602  event is d
month: 10 17602  event is c
**************************** Percentage 62.5
month: 9 17022  event is g
month: 9 17070  event is g
month: 9 17602  event is d
month: 10 17602  event is c
**************************** Percentage 62.5
month: 9 17022  event is d
month: 9 17070  event is g
month: 9 17602  event is d
month: 10 17602  event is c
**************************** Percentage 62.5
month: 9 17022  event is g
month: 9 17022  event is d
month: 9 17612  event is d
month: 10 17602  event is c
**************************** Percentage 62.5
month: 9 17022  event is g
month: 9 17070  event is g
month: 9 17612  event is d
month: 10 17602  event is c
**************************** Percentage 62.5
month: 9 17022  event is d
month: 9 17070  event is g
month: 9 17612  event is d
```

```
month: 10 17602  event is c
**************************** Percentage 62.5
month: 9 17022  event is g
month: 9 17602  event is d
month: 9 17612  event is d
month: 10 17602  event is c
**************************** Percentage 62.5
month: 9 17022  event is d
month: 9 17602  event is d
month: 9 17612  event is d
month: 10 17602  event is c
**************************** Percentage 62.5
month: 9 17070  event is g
month: 9 17602  event is d
month: 9 17612  event is d
month: 10 17602  event is c
```

17022--9-- month —10

```
**************************** Percentage 68.75
month: 10 17070  event is g
month: 11 17070  event is b
**************************** Percentage 68.75
month: 10 17070  event is g
month: 10 17070  event is c
month: 11 17022  event is g
**************************** Percentage 68.75
month: 10 17070  event is g
month: 10 17070  event is c
month: 11 17679  event is b
**************************** Percentage 68.75
month: 10 17070  event is g
month: 10 17070  event is c
month: 11 17680  event is g
**************************** Percentage 68.75
month: 10 17070  event is g
month: 10 17070  event is c
month: 11 17680  event is b
**************************** Percentage 87.5
month: 10 17070  event is g
month: 10 17612  event is c
month: 11 17022  event is g
**************************** Percentage 81.25
month: 10 17070  event is g
month: 10 17612  event is c
month: 11 17679  event is b
**************************** Percentage 87.5
month: 10 17070  event is g
month: 10 17612  event is c
month: 11 17680  event is g
**************************** Percentage 81.25
month: 10 17070  event is g
month: 10 17612  event is c
month: 11 17680  event is b
**************************** Percentage 75.0
month: 10 17070  event is g
month: 10 17612  event is c
month: 11 17070  event is g
```

17022-- month —10

```
*************************** Percentage 68.75
month: 10 17022  event is g
month: 10 17070  event is g
month: 10 17612  event is c
month: 11 17070  event is g
*************************** Percentage 62.5
month: 10 17022  event is g
month: 10 17070  event is g
month: 10 17612  event is c
month: 11 17070  event is b
*************************** Percentage 68.75
month: 10 17022  event is g
month: 10 17070  event is g
month: 10 17612  event is c
month: 11 17602  event is c
```

17022--10-- month —11

```
*************************** Percentage 81.25
month: 11 17070  event is g
month: 12 17022  event is g
*************************** Percentage 68.75
month: 11 17070  event is g
month: 12 17022  event is 6
*************************** Percentage 68.75
month: 11 17070  event is g
month: 12 17022  event is b
*************************** Percentage 75.0
month: 11 17070  event is g
month: 12 17679  event is a
*************************** Percentage 75.0
month: 11 17070  event is g
month: 12 17680  event is a
*************************** Percentage 75.0
month: 11 17070  event is g
month: 12 17070  event is a
*************************** Percentage 68.75
month: 11 17070  event is b
month: 12 17022  event is g
*************************** Percentage 68.75
month: 11 17602  event is c
month: 12 17022  event is g
*************************** Percentage 68.75
month: 11 17602  event is c
month: 12 17679  event is a
*************************** Percentage 68.75
month: 11 17602  event is c
month: 12 17680  event is a
*************************** Percentage 68.75
month: 11 17602  event is c
month: 12 17070  event is a
*************************** Percentage 68.75
month: 11 17070  event is g
month: 12 17612  event is b
```