Available online at www.elixirpublishers.com (Elixir International Journal)

Computer Science and Engineering

Elixir Comp. Sci. & Engg. 62 (2013) 17728-17734

ISSN: 2229-712X

Temporal association rule mining analysis for days temperature

D. P. Rana[1,*], N. J. Mistry[2] and M. M. Raghuwanshi[3]

[1]COED, SVNIT, Surat, India.

[2]CED, SVNIT, Surat, India.

[3]Rajiv Gandhi College of Engineering and Research, Nagpur, India.

ARTICLE INFO

Article history:

Received: 11 July 2013;

Received in revised form:

20 August 2013;

Accepted: 5 September 2013;

Keywords

Association Rules Mining,

Pattern Discovery,

Temporal Association Rules Mining,

Inter Transactional Association Rules Mining.

ABSTRACT

Weather forecasting is a very fundamental application in meteorology and technologically challenging problem. The estimation of temperature values are needed for agricultural, technical and environmental applications. Meteorological dataset has historical data of all weather parameters and the temporal analysis of this dataset can help to mine meaningful knowledge. There are number of techniques available in data mining, but Association Rule Mining is one of the most popular technique to mine large amount of dataset for finding the hidden relationship between various dataset variables values and identifies correlations between them. The scope of this research is to analyze temporal rules generated to predict day to day temperature variation of a specific region Surat, India. To accomplish this, the framework is proposed for prediction of day temperature variation from seasons. From the experiments, achieved higher accuracy compare to other data mining technique and the rules which show how day variations are related. Also prepared the list of parameters which is less in number to help for the prediction instead of all parameters and thus it helps in the reduction of the dataset size.

Introduction

Data mining, also called Knowledge Discovery in Databases (KDD), is the field of discovering novel and potentially useful information from large amounts of data (Han and Kamber, 2012). In contrast to standard statistical methods, data mining techniques search for interesting information without demanding a priori hypotheses, the kind of patterns that can be discovered depending upon the data mining task employed.

Knowledge of meteorological data in a site is essential for number of applications like Rainfall analysis and prediction, temperature prediction, pollution and energy application studies and development. Especially temperature data is used to determine the number of application parameters estimation. An element of weather and the proportion of these elements increases or decreases due to change in climate temperature (Olaiya, 2012).

Meteorologists and weather forecasters make predictions for weather mainly on numerical and statistical models. The simulation conducted often requires intensive computations, involving complex differential equations and computational algorithms. The accuracy is bound by constraints, such as the adoption of incomplete boundary conditions, model assumptions and numerical instabilities, etc. (Olaiya, 2012).

Since the data of everyday weather condition are so huge the weather parameter relations cannot be found easily by directly observation which can be found by data mining techniques. There are number of data mining techniques and methods are available and nowadays used for the prediction and forecasting (Zhang et al., 2004; Huang et al., 2007; Kotsiantis et al., 2007; Kohail et al., 2011; Rana et al., 2011; Sivaramakrishnan et al. 2011; Badhiye et al., 2012).

The meteorological data are huge in amount with number of variables and most of the meteorological variables are related to each other and also vary with each other with respect to time. If the previous year's historical data are used for mining, then system can predict the most accurate value for any weather parameters like temperature.

Here, we need to scan large amount of the dataset for finding the hidden association relationship dataset variables is called association rule mining (Agrawal, 1993) and if we identifies association or correlations between dataset variables with time constraint, this rule mining is referred as temporal association rule mining (Pughazendi et al., 2011; Shahnawaz et al., 2011). From the number of temporal association rule mining types, the inter transaction association rule mining is useful to have temperature relation within

days as the inter transaction association rule mining considers the concept of sliding window and finds the association among the transactions within the sliding window area (Rana et al., 2012). This research is to generate and analyze the temporal association rules for the day to day temperature variation prediction.

Outline

The organization of rest of this report is as follows: Section 2 is discussing the proposed framework for day to day temperature variation prediction, generation and analysis of temporal association rules. It is also providing the results of prediction framework and the performance evaluation of the proposed system. Section 3 is concluding the research work and showing the future plans.

Proposed Framework

Prediction of the future values of temperature by analyzing weather data is one of the important parts which can be helpful to the society as well as to the economy (Olaiya, 2012). In the

Tele:

E-mail addresses: dpr@coed.svnit.ac.in

literature survey it is found that that most of the meteorological data based prediction techniques and methods are based on the statistical or widely used data mining techniques like Clustering, Classification, Regression analysis, Decision Tree etc. but the limited usage of the Association Rule Mining and upto some extent the Temporal Association Rule Mining. Most of the meteorological variables are related to each other and also vary with each other with respect to time. So if we use the previous historical data for prediction then we can predict the most accurate value for any weather parameters. This section is discussing the proposed framework with detailed study of each step for day temperature prediction based on the temporal association rules mining and analysis of generated temporal rules.

As season period is detected, here proposed framework of daily associated pattern of temperature (weather parameter) from season period is derived using inter transactional association rule mining, which is as shown in following Figure 1. Different stages of this framework are as follows:

Dataset Collection

Meteorological data of past 4 year period (2008-2011) for our local place Surat is collected from http://www.wunderground.com. All the parameter contains an average values in dataset observations. The meteorological parameters collected are Temperature, Humidity, Sea Level Pressure, Wind Speed, Precipitation and Visibility.

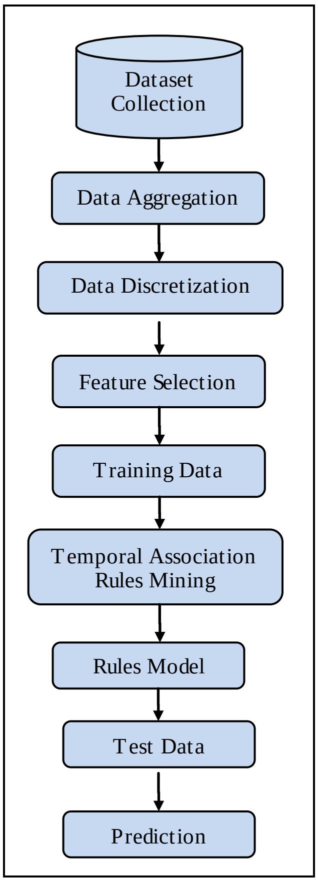

Figure 1: Framework for Day Temperature

Data Aggregation

Our aim is to predict the day to day temperature variation from historical temperature data. For the timely data analysis, from past four year historical temperature data, first the monthly average temperature graph is plotted and analyzed that it has almost same pattern repeated in all four year and from the detailed analysis of this data, we obtain the variation pattern for the particular period of time. According to (Attri et al., 2010) and Climate of India from Wikipedia (http://www.wikipedia.org), it is known to us that the characteristics of these weather parameters are vary according to different time period which is known as season. The weather parameters exhibit almost similar season patterns during all years. From this monthly average temperature graph the one year can be divided into three different seasons 16^{th} February to 15^{th} June period as the Summer season, 16^{th} June to 15^{th} October as the Monsoon and 16^{th} October to next year 15^{th} February as the Winter season. Based on the analyses the four year period (2008-2011) meteorological yearly dataset is separated into seasonal dataset. The data are separated as three years (2008-09, 2009-10, 2010-11) period seasonal data and also separate that yearly data in three different season wise data. By this separation process there are three years three different season data. For example there are three years different monsoon season data, 16^{th} June – 15^{th} October period observations data for the years (2008-09, 2009-10, 2010-11) then that three years monsoon season data are aggregate as one separate dataset. These aggregation procedures are repeated for the summer and the winter season datasets.

Data Discretization

Dataset of meteorological parameters and all parameters are Numerical and the association rule mining is used to mine the data, so first the dataset is converted into the nominal or categorical dataset because the use of categorical dataset is preferable for the association rule mining process. Discretization process transforms a continuous attribute's value into a finite number of intervals. For discretization of the meteorological dataset one class attribute is added into the dataset. The class value represents the day to day variation of the temperature value to the next day temperature value. The class value depends on the next day temperature value whether it is increasing, decreasing or remain constant and by how much degree. An appropriate class label is given to all dataset observations according to the day to day variation of the temperature values. If the differences are 0, +1, -1, +2, -2, +3, -3 and so on, then it's appropriate class values are A, B, b, C, c, D, d and so on respectively.

The CAIM algorithm is used which is the supervised discretizaton type algorithm (Kurgan et al., 2003; Cios, 2004). Supervised algorithm discretizes attributes into the smallest number of intervals by taking into account the interdependence between class labels and the attribute values. Appropriate labels are given to all the attributes according to discretization.

Feature Selection

Since the historical data of everyday weather condition are so huge and have the number of parameters. That many relations between all attributes cannot be found easily by direct observation. And also due to the need and limitation caused by mega transaction, we have to use the feature selection methods to select the most correlated attributes or features or parameters for temperature prediction from six meteorological parameters.

From the literature survey it is found that information gain (IG) is among the most effective method of feature selection and widely used in the field of data mining. Information Gain is defined as the amount of information provided by the feature items or attributes. Information gain is calculated by how much of a particular attribute can be used for classification of dataset observations (Azhagusundari et al., 2013). The feature selection is done using Information Gain attribute evaluator method with the Ranker search method. The ranker search method gives the priority ranks to attributes in which sequence attributes are selected by the attribute evaluator method (Novakovic et al., 2011). If the rank of a particular attribute is greater than the other attribute then it has been given first priority to be chosen from all other attributes.

Based on this priority we can say that the attribute with the highest ranking is the most effective attribute for the dataset of day to day temperature variation. From attributes priority we can choose the most correlated minimum number of attributes for further processing of the temperature prediction. By doing this we are reducing our feature space, making accurate and speeding up our prediction work.

Table 1: Attribute Selection Analysis Result

Information Gain					
Monsoon		**Summer**		**Winter**	
Att.	**Ranking**	**Att.**	**Ranking**	**Att.**	**Ranking**
T	0.389	T	0.289	T	0.308
Precip	0.227	H	0.223	SLP	0.170
H	0.211	WS	0.205	WS	0.169
SLP	0.185	SLP	0.163	H	0.163
WS	0.183	V	0.134	V	0.149
V	0.151	Percip	0.069	Percip	0.082

As shown in Table 1, feature selection and ranking methods do not give information about the removal of any attribute. So, here, we are considering one by one parameter with temperature as input for the further process.

The dataset prepared after feature selection is used for temporal association rule mining, but before it the magatransactions set is generated by user defined window size. The Figure 2 shows sample of generated megatransactions from input dataset with window size 5.

Training Data and Testing Data

The megatransaction dataset is split into two parts using the percentage split method. Take one part as training dataset and another part as testing dataset.

The training dataset is used to train the system and the testing dataset is used to test the system. The dataset is a percentage split using two different methods in different proportions.

First, the random selection method is used. In which the observations are randomly selected from the whole dataset in 50%, 60%, 70% and 75% data as Training datasets and remaining 50%, 40%, 30% and 25 % data as testing datasets. For example if there is a monsoon season dataset then it contains total 366 observations. From those 366 observations 50% or 183 observations are randomly selected as training dataset and other 183 observations as testing datasets.

Second, the yearwise data selection method is used. For example if Monsoon dataset is used and it has total three years monsoon season data. So in year wise data selection method the same proportion of data is selected from three different years' monsoon season data. Monsoon dataset contains total 366 observations. So one year monsoon data have total 122 observations. Therefore from all three years monsoon data the same proportion of data is selected as training data and remaining as testing data.

Temporal Association Rule Mining

The temporal associations rules are generated using inter transaction association rule mining method - Apriori from the megatransaction training dataset for the day's temperature prediction task. The rules are generated where any of the parameters are appearing as antecedent and consequent. But, we need only temperature as consequent and useful rules for the user specified day prediction. This will be done with the help of Rules Model as described in next section.

Rules Model

The Rule Model is one of the most important phases of our framework. A number of rules are generated from inter transaction association rule mining. From that many numbers of the rules, a small number of rules are important rules for day to day temperature prediction purpose. And also the generated rules cannot be directly applied to prediction work, first the rules need to filter based on filter criteria to remove the unwanted and useless rules. To do this filtering consider only that rules that have 1) single or multiple temperature value on left side and only one temperature value on right side of rules. For example, rules like 3T1→5T2 or 5T1 6T2→6T3. 2) The combination of temperature values with other attribute's value at the left side and only single temperature value on right side of the rules. For example, rules like 7SP3 6T3 →6T4.

By applying this filtering, system is reducing the number of rules. The Figure 3 shows the filtered rules from the all generated association rules from training megatransaction dataset with minimum support 2%, minimum confidence 14% and maximum window size for megatransaction is 5. It shows only that many rules which satisfied the filter criteria discussed.

Rule 0: 3T3 → 3T4 Supp. : 0.02 Conf. : 0.21
Rule 1: 3T1 → 3T2 Supp. : 0.02 Conf. : 0.19
Rule 2: 7T0 → 7T1 Supp. : 0.10 Conf. : 0.51
Rule 3: 6T2 → 6T3 Supp. : 0.06 Conf. : 0.35
Rule 4: 6T0 → 6T1 Supp. : 0.06 Conf. : 0.36
Rule 5: 5T1 → 5T2 Supp. : 0.2 0 Conf. : 0.53
Rule 6: 10H1,3T1 → 3T2 Supp. : 0.02 Conf. : 0.31
Rule 7: 10H2,5T0 → 5T3 Supp. : 0.02 Conf. : 0.40
Rule 8: 5SP0,6T0 → 6T1 Supp. : 0.02 Conf. : 0.50
Rule 9: 5SP3,6T3 → 6T4 Supp. : 0.02 Conf. : 0.42

Figure 3: Rules After Rule Filtering

Here the generated rules are not useful for prediction of all day's temperature values. We have rules like T0→T1, T0→T4, T1→T2, T2→T3, T3→T4,.., etc. These rules are applied to predict the temperature value for a particular day. For this reason, first all the filtered rules are converted based on the day difference between both side's temperature values in the rules. The rules generalization process is applied after rules conversion to remove duplicate rules.

For day to day temperature prediction system based on the user given maximum number of days to predict and based on that day value rules conversion process converts the rules in form of day to day prediction rules or day to day variation rules. The Figure 4 shows the converted rules.

From rule conversion process we get the day to day prediction rules but with duplicate rules. So the rules generalization process is applied to convert rules and to remove duplication of the rules. In the rule generalization process, the same rules are removed and keeping the single rule with higher support and confidence value. As shown in Figure 4 with bold style, rules 0-1, 3-4 and 8-9 are duplicate rules. These rules appear more than one time after rule conversion process. The Figure 5 shows the generalized rules after removing duplicate rules and keeping high support and confidence rule. The Figure 6 shows the rules with actual discrete interval values for all parameters instead of parameters labels.

After rule generalization we have final generalized rules and are ready to apply for prediction task. These rules apply on the test dataset to predict the final day temperature value and used to check the prediction accuracy of the proposed framework.

Rule 0:	**3T3 → 3T4**	**Supp.=0.02 Conf.=0.21**
Rule 1:	**3T3 → 3T4**	**Supp.=0.02 Conf.=0.19**
Rule 2:	7T3 → 7T4	Support=0.09 Conf.=0.51
Rule 3:	**6T3 → 6T4**	**Supp.=0.06 Conf.=0.35**
Rule 4:	**6T3 → 6T4**	**Supp.=0.06 Conf.=0.36**
Rule 5:	5T3 → 5T4	Supp.=0.18 Conf.=0.53
Rule 6:	10H3,3T3 → 3T4	Supp.=0.02 Conf.=0.31
Rule 7:	10H3,5T1 → 5T4	Supp.=0.02 Conf.=0.4
Rule 8:	**5SP3,6T3 → 6T4**	**Supp.=0.02 Conf.=0.5**
Rule 9:	**5SP3,6T3 → 6T4**	**Supp.=0.02 Conf.=0.42**

Figure 4: Rules After Rule Conversion Phase

Rule 0:	**3T3 → 3T4**	**Supp.=0.04 Conf.=0.21**
Rule 1:	7T3 → 7T4	Supp.=0.09 Conf.=0.51
Rule 2:	6T3 → 6T4	Supp.=0.12 Conf.=0.36
Rule 3:	**5T3 → 5T4**	**Supp.=0.18 Conf.=0.53**
Rule 4:	10H3,3T3 → 3T4	Supp.=0.02 Conf.=0.31
Rule 5:	10H3,5T1 → 5T4	Supp.=0.02 Conf.=0.4
Rule 6:	**5SP3,6T3 → 6T4**	**Supp.=0.04 Conf.=0.5**

Figure 5: Rules after Rule Generalization Phase

Rule 0:	[25.5-26.5)T3 → [25.5-26.5)T4	Supp.=0.04 Conf.=0.21
Rule 1:	[29.5-30.5)T3 → [29.5-30.5)T4	Supp.=0.09 Conf.=0.51
Rule 2:	[28.5-29.5)T3 → [28.5-29.5)T4	Supp.=0.12 Conf.=0.36
Rule 3:	[27.5-28.5)T3 → [27.5-28.5)T4	Supp.=0.18 Conf.=0.53
Rule 4:	[90.5-96.0]H3, [25.5-26.5)T3 → [25.5-26.5)T4	Supp.=0.02 Conf.=0.31
Rule 5:	[90.5-96.0]H3, [27.5-28.5)T1 → [27.5-28.5)T4	Supp.=0.02 Conf.=0.4
Rule 6:	[1003.5-1005.5)SP3, [28.5-29.5)T3 → [28.5-29.5)T4	Supp.=0.04 Conf.=0.5

Figure 6: Rules with Actual Discrete Interval Values

Now to use those final generalized rules for the prediction the matching is done according to day difference as follows:

- To predict the 5th day temperature value, need of rules like T0→T4, T1→T4, T2→T4, and T3→T4, based on day difference 4, 3, 2 and 1 respectively.
- To predict the 4th day temperature value, need rules like T0→T3, T1→T3, and T2→T3, based on the day difference 3, 2 and 1 respectively.
- To predict the 3rd day temperature value, need rules like T0→T2 and T1→T2, based on the day difference 2 and 1 respectively.
- To predict the 2nd or next day temperature value, need rules like T0→T1, based on the one day difference.

Prediction

For matching and prediction the final generalized association rules are applied one by one onto the test dataset and find the final prediction accuracy for the prediction process. If predicted range values by rules matching with the test data value, then the system takes that predicted temperature value as correctly classified value as per the test dataset.

After applying all the generated rules on the test dataset we have number of correctly classified predicted values from the total number of test data observation. From the number of classified predicted values the accuracy is calculated against the particular test dataset in terms of percentages. For example, there is summer dataset containing total 360 observations and 25% or 90 observations are selected as test data and remaining 75% or 270 as training data. Now, from 90 data 40 data are matched with predicted values. So in this case the prediction accuracy of the system is the percentage ratio of correctly classified observation by total number of data observations. The accuracy of prediction is 45% for this particular example.

Here, the matching and prediction process for one particular test dataset observation is explained with how the system matching all the rules one by one with test data and predicting the future temperature value. For example, we have one generated final rules like 5SP, 5T3→5T4. This rule will be applied to test data observation and predicting range interval 5T4 for the 4th day, if it contains the left part of the rule say 5SP, 5T3. If single rule match is there, then for that observation that predicted value will be final. But, if predicted values by rules application are different, then count the occurrences of all predicted values. And finally the maximum counted value is taken as the final predicted value of all rules for one observation.

Rules analysis

This section presents some useful analysis for temperature prediction for three different seasons. Here the analysis is conducted using three years combine dataset and three year season wise combine datasets. The temporal association rules are generated first from the three years combine dataset and second from the three years different season wise combine datasets. From these generated rules we have explained some analysis points useful for future work.

Rule analysis 1

The system generates rules like 5T3 → 5T4 for monsoon season, 5T3 → 5T4 for the summer season, 10T3 → 10T4 for winter season and 7T3 → 7T4 from three yearly dataset. The Tables 3.7, 3.8 and 3.9 shows the actual interval ranges for attribute labels.

The actual rules like [27.5-28.5) T3 → [27.5-28.5) T4, [27.5-28.5) T3 → [27.5-28.5) T4, [27.5-28.5) T3 → [27.5-28.5) T4. For three year data set it is like [27.5-28.5) T3 → [27.5-28.5) T4. The actual ranges are same for all these rules. So these rules are considered as same rules but generated by separate three year season wise data and presenting the season characteristics and three years data.

From this very important analysis is derived from seasonal rules. Here, the rule from Summer season, range label is 5 which is just above the average temperature of season, for Monson, range label 5 is the average temperature of season and for Winter, range label 10 is the upper limit of the season. And all these rules associate the same range of temperature to very next day temperature.

The system does not know that the particular temperature value is specific season's temperature value. So the association rule mining algorithm associate 3 different season's same range temperature interval values in one rule only, while using 3 year season wise data the algorithm associate only particular season's temperature values in one rule and they can be in lower/mid/upper temperature range.

From this analysis we can say that it is better to use season wise aggregated dataset in place of using 3 years combine dataset.

```
7T0,8WS0,3H0,4SP0,1P0,7T1,8WS1,3H1,5SP1,1P1,7T2,7WS2,3H2,4SP2,1P2,7T3,8WS3,3H3,3SP3,1P3,7T4,8WS4,3H4,3SP4,1P4
7T0,8WS0,3H0,5SP0,1P0,7T1,7WS1,3H1,4SP1,1P1,7T2,8WS2,3H2,3SP2,1P2,7T3,8WS3,3H3,3SP3,1P3,7T4,8WS4,3H4,3SP4,1P4
7T0,7WS0,3H0,4SP0,1P0,7T1,8WS1,3H1,3SP1,1P1,7T2,8WS2,3H2,3SP2,1P2,7T3,8WS3,3H3,3SP3,1P3,7T4,8WS4,3H4,3SP4,1P4
7T0,8WS0,3H0,3SP0,1P0,7T1,8WS1,3H1,3SP1,1P1,7T2,8WS2,3H2,3SP2,1P2,7T3,8WS3,3H3,3SP3,1P3,7T4,8WS4,3H4,3SP4,1P4
7T0,8WS0,3H0,3SP0,1P0,7T1,8WS1,3H1,3SP1,1P1,7T2,8WS2,3H2,3SP2,1P2,7T3,8WS3,3H3,3SP3,1P3,7T4,9WS4,3H4,3SP4,1P4
7T0,8WS0,3H0,3SP0,1P0,7T1,8WS1,3H1,3SP1,1P1,7T2,8WS2,3H2,3SP2,1P2,7T3,9WS3,3H3,3SP3,1P3,7T4,8WS4,3H4,4SP4,1P4
7T0,8WS0,3H0,3SP0,1P0,7T1,8WS1,3H1,3SP1,1P1,7T2,9WS2,3H2,3SP2,1P2,7T3,8WS3,3H3,4SP3,1P3,7T4,7WS4,3H4,5SP4,1P4
7T0,8WS0,3H0,3SP0,1P0,7T1,9WS1,3H1,3SP1,1P1,7T2,8WS2,3H2,4SP2,1P2,7T3,7WS3,3H3,5SP3,1P3,7T4,6WS4,4H4,5SP4,1P4
7T0,9WS0,3H0,3SP0,1P0,7T1,8WS1,3H1,4SP1,1P1,7T2,7WS2,3H2,5SP2,1P2,7T3,6WS3,4H3,5SP3,1P3,7T4,6WS4,4H4,5SP4,1P4
7T0,8WS0,3H0,4SP0,1P0,7T1,7WS1,3H1,5SP1,1P1,7T2,6WS2,4H2,5SP2,1P2,7T3,6WS3,4H3,5SP3,1P3,7T4,6WS4,4H4,5SP4,1P4
7T0,7WS0,3H0,5SP0,1P0,7T1,6WS1,4H1,5SP1,1P1,7T2,6WS2,4H2,5SP2,1P2,7T3,6WS3,4H3,5SP3,1P3,7T4,6WS4,4H4,5SP4,1P4
7T0,6WS0,4H0,5SP0,1P0,7T1,6WS1,4H1,5SP1,1P1,7T2,6WS2,4H2,5SP2,1P2,7T3,6WS3,4H3,5SP3,1P3,7T4,7WS4,3H4,5SP4,1P4
7T0,6WS0,4H0,5SP0,1P0,7T1,6WS1,4H1,5SP1,1P1,7T2,6WS2,4H2,5SP2,1P2,7T3,7WS3,3H3,5SP3,1P3,7T4,6WS4,3H4,5SP4,1P4
:
:
:
```

Figure 2: Screenshot of Megatransaction Generation

Table 3: Accuracy Measurement for Monsoon Season

Attributes Used	Data (%)		Accuracy (%) from Proposed System		Accuracy (%) from Classification	
	Training	Testing	Yearly	Random	Yearly	Random
Temperature	50	50	50	39	16	21
	60	40	43	42	20	19
	70	30	40	45	20	25
	75	25	51	45	16	34
Temperature +Precipitation	50	50	51	42	20	20
	60	40	40	42	20	22
	70	30	44	44	16	23
	75	25	51	47	22	23
Temperature +Precipitation +Humidity	50	50	43	37	21	19
	60	40	42	37	18	21
	70	30	44	43	15	24
	75	25	51	43	20	24
ALL	50	50	40	43	21	23
	60	40	44	44	22	22
	70	30	44	45	22	23
	75	25	50	45	23	19

Table 4: Accuracy Measurement for Summer Season

Attributes Used	Data (%)		Accuracy (%) from Proposed System		Accuracy (%) from Classification	
	Training	Testing	Yearly	Random	Yearly	Random
Temperature	50	50	40	45	14	28
	60	40	44	40	12	25
	70	30	42	45	38	27
	75	25	45	43	16	24
Temperature +Humidity	50	50	40	41	14	21
	60	40	43	41	14	27
	70	30	33	40	35	22
	75	25	35	42	17	23
Temperature +Humidity +Wind Speed	50	50	40	40	14	22
	60	40	45	42	14	16
	70	30	45	45	34	24
	75	25	43	47	18	18
ALL	50	50	37	37	23	13
	60	40	45	37	36	28
	70	30	40	39	31	35
	75	25	37	40	27	21

Table 5: Accuracy Measurement for Winter Season

Attributes Used	Data (%)		Accuracy (%) from Proposed System		Accuracy (%) from Classification	
	Training	Testing	Yearly	Random	Yearly	Random
Temperature	50	50	30	32	20	25
	60	40	27	31	20	25
	70	30	28	32	15	20
	75	25	35	32	23	23
Temperature +SLP	50	50	30	27	20	25
	60	40	25	30	21	23
	70	30	28	30	17	19
	75	25	30	30	20	18
Temperature +SLP +Wind Speed	50	50	28	32	26	17
	60	40	28	33	26	19
	70	30	35	34	14	21
	75	25	42	37	18	18
Temperature +SLP +Wind Speed +Humidity	50	50	39	33	26	25
	60	40	39	36	26	19
	70	30	37	36	15	22
	75	25	42	37	18	17
ALL	50	50	38	32	24	21
	60	40	37	36	25	21
	70	30	35	34	14	21
	75	25	42	34	18	15

Because in 3 year dataset the values of different seasons are associated in one rule, so it is not a proper association for day to day temperature prediction.

- Some time the useful rules may not be generated because of improper associations.

Rule analysis 2

Table 3.11 shows some final generalized rules generated by three different year season wise datasets. The rules in R1 row for different seasons shows that the actual temperature range and predicted temperature ranges are same for each season. The rules in R2 row indicate for summer season that the final predicted temperature range is increase in upper range. For winter season the predicted value range is decrease in lower range and for monsoon season it remains same.

Table 2: Sample Season Wise Rules

Rules	Summer	Winter	Monsoon
R1	6T3==> 6T4	5T3 ==> 5T4	3T3 ==> 3T4
R2	7T3 ==> 8T4	8T3 ==> 7T4	7T3 ==> 7T4
R3	7T1,8T2==>8T4	5T1 5T2 ==> 5T4	7T1,7T2 =>7T4

The rules in raw R3 indicate that if first day temperature interval range is increased in the upper range at 2nd day then the predicted temperature range is same as 2nd day range for the summer season. For winter and monsoon season it remains same.

In Monsoon, Summer and Winter season rules predicted value are from an upper or lower range of intervals, because the interval ranges for these seasons are very small, while for three year data the predicted value are in the same range of intervals, because the interval ranges are bigger than the season wise interval ranges.

Performance analysis

All the experiments were conducted on a Dell-PC, Intel core I5 processor, 2.93 GHz, with 4 GB of RAM running on Windows 7 operating system. Java code is created for the analysis in Eclipse Juno SDK Version 4.2.1. The accuracy of the prediction system is compared with the classification technique on megatransactions. From the weka data mining tool (http://www.cs.waikato.ac.nz/ml/weka/) used J48 algorithm which is the extended version of C4.5 classification algorithm.

Feature selection priority ranking does not provide the information about removal of any parameter. And if we consider all these parameters together, it will increase the size of megatransactions. For each season, first, the accuracy measured using only temperature attribute data then as discussed based on the feature selection priority ranking one by one attribute is considered with each time into the input dataset.

The Table 3 shows the measurements of accuracy values for the monsoon season dataset for temperature variation using only temperature data, then considered precipitation with temperature and then with temperature, precipitation and humidity attributes data and finally using all attribute data. From the experimental analysis of these accuracy values it is found that temperature prediction using the association of only temperature attribute is not a preferable solution. But, we achieved little bit higher accuracy is achieved with Temperature and Precipitation attributes, it will also help in reduction in the size of the inter transactions data.

The Table 4 shows the measurements of accuracy values for the summer season dataset and it is found that with the consideration of Humidity with Temperature, the accuracy is decreasing, compared with the accuracy from only Temperature data. But, with one more attribute Wind Speed together with Temperature and Humidity data, the prediction accuracy value is increased, so these three attributes to consider for summer season.

The Table 5 shows the measurements of accuracy values for the winter season dataset and it is found that if considers Sea level Pressure, Wind Speed and Humidity with Temperature, the accuracy is increasing compare with all other data.

And moreover in all season, for all the cases the accuracy of the proposed prediction system compare to the classification is higher.

Conclusion and Future Work

Temporal association rule mining helps to mine the rules from time stamped data to derive more useful knowledge from the history of meteorological data. Our plan is for the automated system of day to day temperature variation prediction from the historical data using temporal association rule mining.

We have proposed prediction framework useful for temperature prediction analysis on season wise dataset. Based on the performance analysis and experimental results we have observed that 1) Only particular weather parameters are sufficient to predict the day to day temperature from season. 2) Our proposed approach is predicting the temperature with higher accuracy compare to classification algorithm. 3) The proposed approach is simple not very complex. There is no requirement of intensive computations, involving complex differential equations and computational algorithms used for prediction.

From temporal rule analysis to predict the temperature value, we can extend the work further as followings: 1) in the direction to improve the prediction accuracy of our proposed system for temperature prediction. Using data preprocessing techniques like data smoothing, outlier removing. 2) Here, we have predicted the future value for the day from inter transactional rules, where the maximum window size indicates the particular day. It has to be extended for week or month or year by using weekly, monthly and yearly data. 3) To predict the severe weather situation by preparing severe weather condition data for past years. 4) To analyze this prediction system with parameters other than meteorological like solar radiation, cloud cover, latitude, altitude, distance from the sea, phase shift of the earth etc. for the feasibility of improvement.

References

- Agrawal, R.; Imielinksi, T. and Swami, A. (1993), "Mining association rules between sets of items in large database", *ACM SIGMOD Conference on Management of Data*, p. 207–216.
- Attri, S. D. and Tyagi, A. (2010) "Climate profile of India", *Indian Network of Climate Change Assessment National Communication-II,* IMD Delhi.
- Azhagusundari, B. and Thanamani, A. S. (2013), "Feature Selection based on Information Gain", *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, Vol. 2, Iss. 2.
- Badhiye, S. S.; Wakode, B. V. and Chatur, P. N (2012), "Analysis of Temperature and Humidity Data for Future value prediction", *International Journal of Computer Science and Information Technologies, (IJCSIT)*, Vol. 3, Iss. 1, p. 3012 – 3014.
- Cios, L. A. K. (2004), "CAIM Discretization Algorithm", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 16, Iss. 2, p. 145-153.
- Han, J. and Kamber, M. (2012), "Data Mining: Concepts and Techniques", Morgan Kaufmann Publishers.
- Huang, Y. P.; Kao, L. J. and Sandnes, F. E. (2007), "Predicting Ocean Salinity and Temperature Variations Using Data Mining and Fuzzy Inference", *International Journal of Fuzzy Systems*, Vol. 9, Iss. 3.
- Kohail, S. N. and El-Halees, A. M. (2011), "Implementation of Data Mining Techniques for Meteorological Data Analysis", *International Journal of Information and Communication Technology Research (IJICT)*, Vol. 1, Iss. 3.
- Kotsiantis, S.; Kostoulas, A.; Lykoudis, S.; Argiriou, A. and Menagias, K. (2007), "A Hybrid Data Mining Technique for Estimating Mean Daily Temperature Values", *International Journal of Information and Communication Technology Research (IJICT)*, Vol. 1, Iss. 5, p. 54-59.
- Kurgan, L. and Cios, K. (2003), "Fast Class-Attribute Interdependence Maximization (CAIM) Discretization Algorithm", *Proc. of International Conference on Machine Learning and Applications - ICMLA*, p. 30-36.
- Novakovic, J.; Strbac, P. And Bulatovic, D. (2011), "Toward Optimal Feature Selection Using Ranking Methods and Classification Algorithms", *Yugoslav Journal of Operations Research*, Vol. 1, p. 119-135.
- Olaiya, F. (2012), "Application of Data Mining Techniques in Weather Prediction and Climate Change Studies", *International Journal of Information Engineering and Electronic Business*, Vol. 1, p. 51–59.
- Pughazendi, N. and Punithavalli, M. (2011), "Discovering Patterns from Temporal Databases Using Temporal Association Rule", *Global Journal of Computer Science and Technology*, ISSN: 0975-4350, Vol. 11, Iss. 14, p. 1–5.
- Rana, D. P.; Mistry, N. J. and Raghuwanshi, M. M. (2011), "Association Rule Mining for Environmental Data: A Study", Intl. Congress of Environmental Research (ICER-11), India.
- Rana, D. P.; Mistry, N. J. and Raghuwanshi, M. M. (2012), "Inter Transactional Association Rule Mining using Boolean Matrix", *International Journal of Research in Computer and Communication Technology (IJRCCT)*, Vol. 1, Iss. 2, p. 26-33.
- Shahnawaz, M.; Ranjan, A. and Danish, M. (2011), "Temporal Data Mining: An Overview", *International Journal of Engineering and Advanced Technology (IJEAT)*, ISSN: 2249 – 8958, Vol. 1, Iss. 1, p. 20–24.
- Sivaramakrishnan, T. R. and Meganathan, S. (2011) "Association Rule Mining and Classifier Approach for Quantitative Spot Rainfall Prediction", *Journal of Theoretical and Applied Information Technology*, Vol. 34, Iss. 2.