

Discovering Sequential Association Rules with Constraints and Time Lags in Multiple Sequences^{*}

Sherri K. Harms¹, Jitender Deogun², and Tsegaye Tadesse³

¹ Department of Computer Science and Information Systems
University of Nebraska-Kearney, Kearney NE 68849 harmssk@unk.edu,
<http://faculty.unk.edu/h/harmssk>

² Department of Computer Science and Engineering
University of Nebraska-Lincoln, Lincoln NE 68588-0115

³ National Drought Mitigation Center
University of Nebraska-Lincoln, Lincoln NE 68588

Abstract. We present *MOWCATL*, an efficient method for mining frequent sequential association rules from multiple sequential data sets with a time lag between the occurrence of an antecedent sequence and the corresponding consequent sequence. This approach finds patterns in one or more sequences that precede the occurrence of patterns in other sequences, with respect to user-specified constraints. In addition to the traditional frequency and support constraints in sequential data mining, this approach uses separate antecedent and consequent inclusion constraints. Moreover, separate antecedent and consequent maximum window widths are used to specify the antecedent and consequent patterns that are separated by the maximum time lag.

We use multiple time series drought risk management data to show that our approach can be effectively employed in real-life problems. The experimental results validate the superior performance of our method for efficiently finding relationships between global climatic episodes and local drought conditions. We also compare our new approach to existing methods and show how they complement each other to discover associations in a drought risk management decision support system.

1 Introduction

Discovering association rules in sequences is an important data-mining problem that is useful in many scientific and commercial domains. Predicting events and identifying sequential rules that are inherent in the data help domain experts to learn from past data and make informed decisions for the future. Several different approaches have been investigated for sequential data mining [1], [2], [3], [4], [5]. Algorithms for discovering associations in sequential data [2], and episodal

^{*} This research was supported in part by NSF Digital Government Grant No. EIA-0091530 and NSF EPSCOR, Grant No. EPS-0091900.

associations [1], [3] use all frequent episodes. The entire set of association rules is produced and significance criterion such as J-measure for rule ranking are used to determine the valuable rules [2]. An approach that uses temporal constraints on transactional sequences was presented in [5]. Our earlier methods, *Gen-FCE* and *Gen-REAR*, use inclusion constraints with a sliding window approach on event sequences to find the frequent closed episodes and then generate the representative episodal association rules from those episodes. We propose a generalized notion of episodes where the antecedent and consequent patterns are separated by a time lag and may consist of events from multiple sequences.

In this paper, we present a new approach that uses Minimal Occurrences With Constraints And Time Lags (*MOWCATL*), to find relationships between sequences in the multiple data sets. In addition to the traditional frequency and support constraints in sequential data mining, *MOWCATL* uses separate antecedent and consequent inclusion constraints, along with separate antecedent and consequent maximum window widths, to specify the antecedent and consequent patterns that are separated by a maximum time lag. The *MINEPI* algorithm was the first approach to find minimal occurrences of episodes [3].

Our approach is well suited for sequential data mining problems that have groupings of events that occur close together, but occur relatively infrequently over the entire dataset. They are also well suited for problems that have periodic occurrences when the signature of one or more sequences is present in other sequences, even when the multiple sequences are not globally correlated. The analysis techniques developed in this work facilitate the evaluation of the temporal associations between episodes of events and the incorporation of this knowledge into decision support systems. We show how our new approach complements the existing approaches to address the drought risk management problem.

2 Events and Episodes

For mining, sequential datasets are normalized and discretized to form subsequences using a *sliding window* [2]. With a sliding window of size δ , every normalized time stamp value at time t is used to compute each of the new sequence values $y_{t-\delta/2}$ to $y_{t+\delta/2}$. Thus, the dataset is divided into segments, each of size δ . The discretized version of the time series is obtained by using a clustering algorithm and a suitable similarity measure [2]. We consider each cluster identifier as a single *event type*, and the set of cluster labels as the *class of events* E . The new version of the time series is called an event sequence. Formally, an *event sequence* is a triple (t_B, t_D, \mathcal{S}) where t_B is the beginning time, t_D is the ending time, and \mathcal{S} is a finite, time-ordered sequence of events [3], [6]. That is, $\mathcal{S} = (e_{t_B}, e_{t_{B+1p}}, e_{t_{B+2p}}, \dots, e_{t_{B+dp}} = e_{t_D})$, where p is the step size between events, d is the total number of steps in the time interval from $[t_B, t_D]$, and $D = B + dp$. Each e_{t_i} is a member of a class of events E , and $t_i \leq t_{i+1}$ for all $i = B, \dots, D - 1p$. A sequence of events \mathcal{S} includes events from a single class of events E .

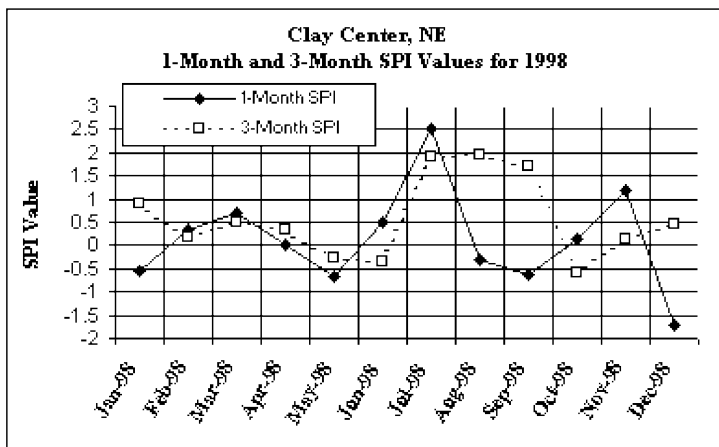


Fig. 1. Example multiple event sequences

Example 1. Consider the event sequences of 1-month and 3-month Standardized Precipitation Index (SPI) values from Clay Center, Nebraska from January to December 1998 shown in Figure 1. SPI values show rainfall deviation from normal for a given location at a given time [7]. For this application, a sliding window width of 1 month was used, and the data was clustered into 7 clusters: A. Extremely Dry ($SPIvalue \leq -2.0$), B. Severely Dry ($-2.0 < SPIvalue \leq -1.5$), C. Moderately Dry ($-1.50 < SPIvalue \leq -0.5$), D. Normal ($-0.5 < SPIvalue < 0.5$), E. Moderately Wet ($0.5 \leq SPIvalue < 1.5$), F. Severely Wet ($1.5 \leq SPIvalue < 2.0$), and G. Extremely Wet ($SPIvalue \geq 2.0$).

When multiple sequences are used, each data set is normalized and discretized independently. The time granularity is then converted to a single (finest) granularity [1] before the discovery algorithms are applied to the combined sequences.

An episode in an event sequence is a partial order defined on a set of events [3], [6]. It is said to occur in a sequence if events are consistent with the given order, within a given time bound (window width). Formally, an *episode* P is a pair $(V, type)$, where V is a collection of events. An episode is of type *parallel* if no order is specified and of type *serial* if the events of the episode have a fixed order. An episode is *injective* if no event type occurs more than once in the episode.

3 The MOWCATL Method

The MOWCATL method shown in Figure 2, finds minimal occurrences of episodes and relationships between them and requires a single database pass as in MINEPI algorithm[3]. Larger episodes are built from smaller episodes by joining episodes with overlapping minimal occurrences, which occur within the

specified window width. However, our approach has additional mechanisms for: (1) constraining the search space during the discovery process, (2) allowing a time lag between the antecedent and consequent of a discovered rule, and (3) working with episodes from across multiple sequences. Our focus is on finding episodal rules where the antecedent episode occurs within a given maximum window width win_a , the consequent episode occurs within a given maximum window width win_c , and the start of the consequent follows the start of the antecedent within a given maximum time lag . This allows us to easily find rules such as “if A and B occur within 3 months, then within 2 months they will be followed by C and D occurring together within 4 months.”

- 1) Generate Antecedent Target Episodes of length 1 ($ATE_{1,B}$);
- 2) Generate Consequent Target Episodes of length 1 ($CTE_{1,B}$);
- 3) Input sequence S , record occurrences of $ATE_{1,B}$ and $CTE_{1,B}$ episodes;
- 4) Prune unsupported episodes from $ATE_{1,B}$ and $CTE_{1,B}$;
- 5) $k = 1$;
- 6) **while** ($ATE_{k,B} \neq \emptyset$) **do**
- 7) Generate Antecedent Target Episodes $ATE_{k+1,B}$ from $ATE_{k,B}$
- 8) Record each minimal occurrence of the episodes less than win_a ;
- 9) Prune the unsupported episodes from $ATE_{k+1,B}$;
- 10) $k++$;
- 11) Repeat or execute in parallel, Steps 5 - 11 for consequent episodes, using $CTE_{k+1,B}$ and win_c ;
- 12) Generate combination episodes CE_B from $ATE_B \times CTE_B$;
- 13) Record the combination’s minimal occurrences that occur within lag ;
- 14) Return the supported lagged episode rules in CE_B that meet the min_conf threshold;

Fig. 2. MOWCATL Algorithm.

Our approach is based on identifying minimal occurrences of episodes along with their time intervals. Given an episode α and an event sequence S , we say that the window $w = [t_s, t_e]$ is a *minimal occurrence* of α in S , if: (1) α occurs in the window w , and (2) α does not occur in any proper subwindow of w . The maximal width of a minimal occurrence for both the antecedent and the consequent are fixed during the process, and will measure the interestingness of the episodes.

The sequence S can be a combination of multiple sequences S_1, S_2, \dots, S_k . An episode can contain events from each of the k sequences. Additionally, *combination events* are created with events from different sequences that occur together at the same timestamp. When finding minimal occurrences, a combination event is considered as a single event. The *support* of an episode α is the number of minimal occurrences of α . An episode α is considered *frequent* if its support meets or exceeds the given minimum support threshold min_sup . After the frequent episodes are found for the antecedent and the consequent independently, we combine the frequent episodes to form an episode rule.

Definition 1. An *episode rule* r is defined as an expression $\alpha[win_a] \Rightarrow_{lag} \beta[win_c]$, where α and β are episodes, and win_a , win_c , and lag are integers.

For each frequent antecedent episode α we join its minimal occurrences with each minimal occurrence of each frequent consequent episode β , as long as the starting time of the minimal occurrence for β is after the starting time of the minimal occurrence of α and no later than the ((starting time of α) + lag). The occurrence of α must end before the occurrence of β ends. The number of events in episodes α and β may differ. The informal interpretation of the rule is that if episode α has a minimal occurrence in the interval $[t_s, t_e)$, with $t_e - t_s \leq win_a$, and β has a minimal occurrence in the interval $[t_r, t_d)$, with $t_d - t_r \leq win_c$, and t_r is in the range $[t_{s+1}, t_{s+lag}]$, and $t_e < t_d$, then the rule r has a minimal occurrence in the interval $[t_s, t_d)$.

The *confidence* of an episode rule $r = \alpha[win_a] \Rightarrow_{lag} \beta[win_c]$ in a sequence S with given windows win_a , win_c , and lag is the conditional probability that β occurs, given that α occurs, under the time constraints specified by the rule. The *support* of the rule is the number of times the rule holds in the database.

Example 2. The *MOWCATL* method generates minimal occurrences and episodal rules shown in Table 1 when applied to event sequences S given in Figure 1, with $win_a = 3$, $min_sup = 2$, $win_c = 3$, $lag = 1$, with the SPI1 sequence as the antecedent and the SPI3 sequence as the consequent for parallel episodes. The events are the cluster labels described in Example 1.

Table 1. Sample *MOWCATL* episodes, minimal occurrences, and rules

Episode/Rule	Minimal occurrences	Support	Confidence
1C	1-1, 5-5, 9-9	3	
1D	2-2, 4-4, 6-6, 8-8, 10-10	5	
1E	3-3, 11-11	2	
3D	2-2, 3-3, 4-4, 5-5, 6-6, 11-11, 12-12	7	
3F	7-7, 8-8, 9-9	3	
1C,1D	1-2, 4-5, 5-6, 8-9, 9-10	5	
1C,1E	1-3, 9-11	2	
1D,1E	2-3, 3-4, 10-11	3	
3D,3F	6-7, 9-11	2	
1C,1D,1E	1-3, 3-5, 9-11	3	
1C,1D $\Rightarrow_{lag=1}$ 3D,3F	(5-6,6-7), (8-9,9-11)	2	.4
1C,1D $\Rightarrow_{lag=1}$ 3D	(1-2,2-2), (4-5,5-5), (5-6,6-6)	3	.6
1D,1E $\Rightarrow_{lag=1}$ 3D	(2-3,3-3), (3-4,4-4), (10-11, 11-11)	3	1
1C $\Rightarrow_{lag=1}$ 3D	(1-1,2-2), (5-5,6-6)	2	.67
1D $\Rightarrow_{lag=1}$ 3D	(2-2,3-3), (4-4,5-5), (10-10, 11-11)	3	.6
1D $\Rightarrow_{lag=1}$ 3F	(6-6,7-7), (8-8,9-9)	2	.4
1E $\Rightarrow_{lag=1}$ 3D	(3-3,4-4), (11-11,12-12)	2	1

4 The *Gen-FCE* and the *Gen-REAR* Methods

Previously, we presented the *Gen-FCE* and *Gen-REAR* methods for the drought risk management problem [8]. *Gen-FCE*, defines a *window* on an event sequence S as an event subsequence $W = \{e_{t_j}, \dots, e_{t_k}\}$, where $t_B \leq t_j$, and $t_k \leq t_D + 1$ as in

the *WINEPI* algorithm[3], [6]. The width of the window W is $width(W) = t_k - t_j$. The set of all windows W on \mathcal{S} , with $width(W) = win$ is denoted as $\mathcal{W}(\mathcal{S}, win)$. The window width is pre-specified. The *frequency* of an episode is defined as the fraction of windows in which the episode occurs. Given an event sequence \mathcal{S} , and a window width win , the frequency of an episode P of a given type in \mathcal{S} is:

$$fr(P, \mathcal{S}, win) = \frac{|\mathbf{w} \in \mathcal{W}(\mathcal{S}, win) : P \text{ occurs in } \mathbf{w}|}{|\mathcal{W}(\mathcal{S}, win)|}$$

Given a frequency threshold min_fr , P is *frequent* if $fr(P, \mathcal{S}, win) \geq min_fr$. Closure of an episode set X , denoted by $closure(X)$, is the smallest closed episode set containing X and is equal to the intersection of all frequent episode sets containing X . *Gen-FCE* generates frequent closed target episodes with respect to a given set of Boolean target constraints B , an event sequence \mathcal{S} , a window width win , an episode type, a minimum frequency min_fr , and a window step size p . We use the set of frequent closed episodes *FCE* produced from the *Gen-FCE* algorithm to generate the *representative episodal association rules* (REAR) that cover the entire set of association rules [9].

Using our techniques on multiple time series while constraining the episodes to a user-specified target set, we can find relationships that occur across the sequences. Once the set of representative association rules is found, the user may formulate queries about the association rules that are covered (or represented) by a certain rule of interest for given support and confidence values. These techniques can be employed in many problem domains, including drought risk management.

5 Drought Risk Management – An Application

Drought affects virtually all US regions and results in significant economic, social, and environmental impacts. According to the National Climatic Data Center, the losses due to drought are more than any other severe weather disaster. Given the complexity of drought, where the impacts from a drought can accumulate gradually over time and vary widely across many sectors, a well-designed decision support system is critical to effectively manage drought response efforts.

This work is part of a Digital Government project at UNL that is developing and integrating new information technologies for improved government services in the USDA Risk Management Agency (RMA) and the Natural Resources Conservation Service. We are in the process of developing an advanced Geospatial Decision Support System (GDSS) to improve the quality and accessibility of drought related data for drought risk management [10]. Our objective is to integrate spatio-temporal knowledge discovery techniques into the GDSS using a combination of data mining techniques applied to geospatial time-series data.

6 Experimental Results and Analysis

Experiments were designed to find relationships between drought episodes at the automated weather station in Clay Center, NE, and other climatic episodes,

from 1949-1999. There is a network of automated weather stations in Nebraska that can serve as long-term reference sites to search for key patterns and link to climatic events. We use historical and current climatology datasets, including 1) Standardized Precipitation Index (SPI) data from the National Drought Mitigation Center (NDMC), 2) Palmer Drought Severity Index (PDSI) from the National Climatic Data Center (NCDC), 3) North Atlantic Oscillation Index (NAO) from the Climatic Research Unit at the University of East Anglia, UK, 4) Pacific Ocean Southern Oscillation Index (SOI) and Multivariate ENSO Index (MEI) available from NOAA's Climate Prediction Center, and 5) Pacific/North American (PNA) Index and Pacific Decadal Oscillation (PDO) Index available from the Joint Institute for the Study of the Atmosphere and Ocean.

The data for the climatic indices are grouped into seven categories, i.e. extremely dry, severely dry, moderately dry, near normal, moderately wet, severely wet, and extremely wet. In our study, the 1-month, 3-month, 6-month, 9-month, 12-month SPI, and the PDSI values are grouped into the same seven categories to show the precipitation intensity relative to normal precipitation for a given location and a given month. The SOI, MEI, NAO, PDO, and PNA categories are based on the standard deviation from the normal and the negative values are considered to show the dry periods.

After normalizing and discretizing each dataset using the seven categories above, we performed experiments to find whether the method discovers interesting rules from the sequences, and whether the method is robust. Several window widths, minimal frequency values, minimal confidence values, and time lag values for both parallel and serial episodes were used. We specified droughts (the three dry categories in each data source) as our target episodes. For *MOWCATL*, we used the global climatic indices (SOI, MEI, NAO, PDO, and PNA) as our antecedent data sets, and the local precipitation indices (SPI1, SPI3, SPI6, SPI9, SPI12, and PDSI) as our consequent data sets. The experiments were ran on a DELL Optiplex GX240 2.0GHz PC with 256 MB main memory, under the Windows 2000 operating system. Algorithms were coded in C++.

Episodes with time lags from *MOWCATL* are useful to the drought risk management problem when trying to predict future local drought risk considering the current and past global weather conditions. Table 2 represent performance statistics for finding frequent drought episodes with various support thresholds using the *MOWCATL* algorithm. *MOWCATL* performs extremely well when finding the drought episodes. At a minimum support of .020 for parallel episodes, the algorithm only needs to look through the 212 candidate drought episodes to find the 109 frequent drought episodes. Whereas, using no constraints it would need to look through 3892 candidate episodes to find 2868 total frequent episodes.

Gen-FCE episodes are useful to the drought risk management problem when considering events that occur together, either with order (serial episodes), or without order specified (parallel episodes). Table 3 represent performance statistics for finding frequent closed drought episodes with various frequency thresholds using the *Gen-FCE* algorithm. As shown, the number of frequent closed

Table 2. Performance characteristics for parallel and serial drought episodes and rules with *MOWCATL*, Clay Center, NE drought monitoring database, $win_a = 4$ months, $win_c = 3$ months, and $lag = 2$ months, and $min_conf = 25\%$.

Min. support	<i>Parallel</i>				<i>Serial</i>			
	Total cand.	Freq. episodes	Distinct rules	Total time (s)	Total cand.	Frequent episodes	Distinct rules	Total time (s)
0.005	930	732	98	3	9598	1125	174	34
0.007	716	575	41	3	6435	621	58	33
0.010	452	288	10	2	4144	275	15	32
0.013	332	192	7	1	3457	168	6	32
0.016	254	142	1	1	2805	109	1	31
0.020	212	109	1	1	2637	83	1	30

episodes decreases rapidly as the frequency threshold increases as expected from the nature of drought.

Tables 2 and 3 also show the number of distinct rules generated for these algorithms. As shown, the number of rules between global climatic drought episodes and local drought at Clay Center, NE decreases rapidly as the frequency and support levels increase. In fact, there was only one parallel drought rule out of 1954 total rules at a 25% confidence level for a support threshold of 0.020 using the *MOWCATL* algorithm.

Examples of how the window widths influence the results are shown in Table 4. The *MOWCATL* algorithm finds a significant number of patterns and relationships for all window widths specified. In general, wider combined window widths win_a , win_c , produce more patterns and relationships, but with less significant meaning. With a 2 month lag in time, the *MOWCATL* algorithm discovers 142 parallel drought episodal rules and 199 serial drought episodal rules, using $win_a = 3$ and $win_c = 3$. *MOWCATL* discovers more relationships at higher confidence values than the *Gen-REAR* approach. These examples indicate that there is a delay in time after global climatic drought indicators are recognized, before local drought conditions occur. This is encouraging, since knowing this time difference will allow drought risk management experts time to plan for the expected local drought conditions.

Table 3. Performance characteristics for parallel and serial drought episodes and rules with *Gen-FCE* and *Gen-REAR*, Clay Center, NE drought monitoring database, window width 4 months and a $min_conf = 25\%$.

Min. freq.	<i>Parallel</i>				<i>Serial</i>			
	Total cand.	Freq. episodes	Distinct rules	Total time (s)	Total cand.	Frequent episodes	Distinct rules	Total time (s)
0.02	1891	93	175	4	3002	327	9	10
0.04	650	265	41	1	1035	139	1	6
0.08	297	68	10	0	494	33	0	1
0.12	154	28	1	0	310	16	0	0
0.16	108	15	1	0	226	13	0	0
0.20	75	10	0	0	160	10	0	0
0.24	51	7	0	0	112	7	0	0

Table 4. Performance characteristics for parallel and serial drought episodes and rules for the Clay Center, NE drought monitoring database, with varying window widths. Parameters include $lag = 2$, $min_sup = 0.005$, $min_fr = 0.02$, and $min_conf = 25\%$.

win or win_a	win_c	<i>MOWCATL</i>				<i>Gen-FCE/Gen-REAR</i>			
		<i>Parallel</i>		<i>Serial</i>		<i>Parallel</i>		<i>Serial</i>	
		Freq. episodes	Distinct rules	Freq. episodes	Distinct rules	Freq. episodes	Distinct rules	Freq. episodes	Distinct rules
1	1	135	44	101	45	40	0	40	0
1	2	319	64	387	55				
1	3	532	72	741	58				
2	1	199	84	212	134	149	4	55	0
2	2	383	127	498	176				
2	3	596	154	852	189				
3	1	252	79	331	140	400	45	183	1
3	2	436	121	596	184				
3	3	649	142	951	199				
4	1	335	60	485	125	930	175	327	9
4	2	519	86	771	164				
4	3	732	98	1125	174				
4	4	1056	104	1596	198				

Finding the appropriate time lag is an iterative process. Using the parameters from Table 2, but decreasing the time lag to one month, reduces the number of rules to 24 parallel drought rules and 62 serial drought rules at a minimal support of 0.005. By increasing the time lag to three months, we get 275 parallel drought rules and 506 serial drought rules. As the time lag increases, more rules are discovered, but again with decreased significant meaning.

Clearly, the results produced by these methods need to be coupled with human interpretation of the rules and an interactive approach to allow for iterative changes in the exploration process. Using our methods, the drought episodes and relationships are provided quickly and without the distractions of the other non-drought data. These are then provided to the drought risk management expert for human interpretation. We provide the user with the J-measure [2] for ranking rules by interestingness, rather than using the confidence value alone. Similarly, our method can be employed in other applications.

7 Conclusion

This paper presents a new approach for generating episodal association rules in multiple data sets. We compared the new approach to the *Gen-FCE* and the *Gen-REAR* approaches, and showed how the new approach complements these techniques in addressing complex real-life problems like drought risk management. As demonstrated by the experiments, our methods efficiently find relationships between climatic episodes and droughts by using constraints, time lags, closures and representative episodal association rules.

Other problem domains could also benefit from this approach, especially when there are groupings of events that occur close together in time, but occur relatively infrequently over the entire dataset. Additional suitable problem domains are when the entire set of multiple time series is not correlated, but there are periodic occurrences when the signature of one sequence is present in other sequences, with possible time delays between the occurrences. The analysis techniques developed in this work facilitate the evaluation of the temporal associations between episodes of events and the incorporation of this knowledge into decision support systems. Currently, there is no commercial product that addresses these types of problems.

For future work, we plan to extend these methods to consider the spatial extent of the relationships. Additionally, we are incorporating these approaches into the advanced geospatial decision support system for drought risk management mentioned above.

References

1. Bettini, C., Wang, X.S., Jajodia, S.: Discovering frequent event patterns with multiple granularities in time sequences. *IEEE Transactions on Knowledge and Data Engineering* **10** (1998) 222–237
2. Das, G., Lin, K.I., Mannila, H., Ranganathan, G., Smyth, P.: Rule discovery from time series. In: *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining [KDD 98]*, New York, NY (1998) 16–22
3. Mannila, H., Toivonen, H., Verkamo, A.I.: Discovery of frequent episodes in event sequences. Technical report, Department of Computer Science, University of Helsinki, Finland (1997) Report C-1997-15.
4. Srikant, R., Vu, Q., Agrawal, R.: Mining association rules with item constraints. In: *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining [KDD97]*. (1997) 67–73
5. Zaki, M.: Sequence mining in categorical domains: Incorporating constraints. In: *Proceedings of the Ninth International Conference on Information and Knowledge Management [CIKM2000]*, Washington D.C., USA (2000) 422–429
6. Mannila, H., Toivonen, H., Verkamo, A.I.: Discovering frequent episodes in sequences. In: *Proceedings of the First International Conference on Knowledge Discovery and Data Mining [KDD 95]*, Montreal, Canada (1995) 210–215
7. McGee, T.B., Doeskin, N.J., Kliet, J.: Drought monitoring with multiple time scales. In: *Proceedings of the 9th Conference on Applied Climatology*, Boston, MA (1995) 233–236 American Meteorological Society.
8. Harms, S.K., Deogun, J., Saquer, J., Tadesse, T.: Discovering representative episodal association rules from event sequences using frequent closed episode sets and event constraints. In: *Proceedings of the 2001 IEEE International Conference on Data Mining*, San Jose, California, USA (2001) 603–606
9. Kryszkiewicz, M.: Fast discovery of representative association rules. In: *Lecture Notes in Artificial Intelligence*. Volume 1424., *Proceedings of RSCTC 98*, Springer-Verlag (1998) 214–221
10. Harms, S.K., Goddard, S., Reichenbach, S.E., Waltman, W.J., Tadesse, T.: Data mining in a geospatial decision support system for drought risk management. In: *Proceedings of the 2001 National Conference on Digital Government Research*, Los Angeles, California, USA (2001) 9–16