Frequent and Significant Episodes in Sequences of Events
Computation of a New Frequency Measure based on Individual Occurrences of the Events

Oscar Quiroga, Joaquim Meléndez and Sergio Herraiz
Institute of Informatics and Applications, University of Girona, 17071 Girona, Catalonia, Spain

Keywords: Data Mining, Event Sequences, Frequent Episodes, Pattern Discovery.

Abstract: Pattern discovery in event sequences is based on the mining of frequent episodes. Patterns are the result of the assessment of frequent episodes using episode rules. However, with a simple search usually a huge number of frequent episodes and rules are found, then, methods to recognise the most significant patterns and to properly measure the frequency of the episodes, are required. In this paper, two new indexes called cohesion and backward-confidence of the episodes are proposed to help in the extraction of significant patterns. Also, two methods to find the maximal number of non-redundant occurrences of serial and parallel episodes are presented. Experimental results demonstrate the compactness of the mining result and the efficiency of our mining algorithms.

1 INTRODUCTION

Frequent episodes mining in sequences of events can be applied in many application domains, as for example in telecommunication networks (Mannila et al., 1997), web access pattern discovery, family protein analysis (Casas-Garriga, 2003), fault prognosis based on logs of a manufacturing plant (Laxman et al., 2007), study of multi-neuronal spike train recordings (Patnaik, 2006) or event tracking problems for news stories (Iwanuma et al., 2005), etc. The starting point are data sets organised as a single long sequence of events where each event is described by its type and its time of occurrence.

Pattern discovery in event sequences includes two main steps: frequent episodes extraction and significant episodes recognition. Frequent episodes extraction is usually a process that starts looking for frequent single events (frequent events) and, following an iterative procedure, candidate episodes are build from frequent episodes found in the previous level. Candidates who reach a minimum frequency threshold are classified as frequent episodes(Agrawal and Srikant, 1995). Several methods have been proposed to cope with particularities of events and episodes, considering for example duration, maximal gap between events, overlapping/non-overlapping of episodes, minimal occurrences, etc. The frequency of an episode i.e., the number of occurrences over a sequence, may vary among different algorithms and results depend on these procedures but also on how events are distributed in the sequence (Gan and Dai, 2010).

Relevance of frequent episodes have to be evaluated to find representative connections between events describing *patterns*. Confidence of the episode rule can be used for that purpose (Agrawal and Srikant, 1994). However, with a simple search usually a huge number of frequent episodes and rules are found. Thus, to extract the most relevant information it is necessary to use auxiliary criteria (Gan and Dai, 2011).

In this paper, an improved frequency measure method and two new indexes to assess the significance of the episodes are suggested. The method locates and counts the largest number of non-redundant occurrences of an episode, and is useful for both, serial and parallel episodes. The new indexes are proposed as complementary criteria to the confidence of the episode rules. The first one, named *cohesion of the episode*, is based on the comparison of the number of serial and parallel occurrences, whereas the second, named *backward-confidence of the episode*, is analogous to the confidence of the episode rule but evaluates the beginning of the episode.

The usefulness of the method and the indexes proposed, are evaluated with synthetic event sequences.

Quiroga O., Meléndez J. and Herraiz S..
Frequent and Significant Episodes in Sequences of Events - Computation of a New Frequency Measure based on Individual Occurrences of the Events.
DOI: 10.5220/0004118003240328
In *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval* (KDIR-2012), pages 324-328
ISBN: 978-989-8565-29-7

2 FMINEVENT

Fminevent is the short name of the *frequency measure method based on individual occurrences of the events* proposed in this paper. The method is described with commonly used terms in literature of sequence data mining.

Event: An event is defined by the pair *(e,t)* where t denotes the occurrence time (timestamp) and e represent the event attributes (one or several) that contain the information useful to characterise the event. Event attributes can be a single label or a vector of continuous/discrete attribute-value pairs defined in a given range or set of predefined values.

Sequence of Events: A sequence of events S is defined as an ordered list of events or a n-tuple $S = \langle(e_1,t_1),(e_2,t_2),...,(e_n,t_n)\rangle$ where $t_i \leq t_{i+1}$ for all $i \in \{1,2,...,n-1\}$. The length of S, $|S|$, is n. In single sequence mining, the events are represented categorically from a finite set of event types.

Episode: An episode α is ordered list of characterised events E of the form $\alpha = \langle a_1,a_2,...,a_m\rangle$ with $a_j \in E$ for all $j = 1,...,m$. The size of α is the number of elements of α that is $|\alpha| = m$. An episode α imposes a constrain on relative order of occurrences of a_j i.e., if the event type a_j occurs before event type a_{j+1} for all $j = 1,...,m-1$ is a *serial episode*. If there are no constraints about the order of their appearances is called *parallel episode* and will be denoted as $\alpha = \langle a_1 \cdot a_2 \cdot ... \cdot a_m\rangle$.

Sub-episode, Super-episode and Maximal Frequent Episode: An episode $\beta = \langle b_1,b_2,...,b_m\rangle$ is a sub-episode of another episode $\alpha = \langle a_1,a_2,...,a_n\rangle$ if there exist $1 \leq i_1 < i_2 < ... < i_m \leq n$ such that $b_j = a_{i_j}$ for all $j = 1,2,\ldots,m$. In this case, α is a super-episode of β. If we have $i_1 = 1, i_2 = 2, \cdots, i_m = m$, then α is called the **forward-extension** super episode of β. If we have $i_1 = n-m+1, i_2 = n-m+2, \cdots, i_m = n$, then α is called the **backward-extension** super episode of β. When an episode do not have a super-episode it is called a **maximal frequent episode**.

Occurrences: An episode $\alpha = \langle a_1,a_2,...,a_m\rangle$ occurs in a sequence of events $S = \langle(e_1,t_1),(e_2,t_2),...,(e_n,t_n)\rangle$ if there is at least one ordered sequence of events $S' = \langle(e_{i_1},t_{i_1}),(e_{i_2},t_{i_2}),...,(e_{i_m},t_{i_m})\rangle$ such that $S' \subseteq S$ and $a_j = e_{i_j}$ for all $j = 1,2,...,m$. Usually an occurrence is denoted as $o = \langle i_1,i_2,...,i_m\rangle$ where $o[j] = i_j$ and $j = 1,2,...,m$.

Non Redundant Occurrences: A set of occurrences of an episode α is called non-redundant if for any two occurrences $o = \langle i_1,i_2,...,i_m\rangle$ and $o' = \langle i'_1,i'_2,...,i'_m\rangle$ no event occurs simultaneously in both, i. e., $e_{i_j} \neq e_{i'_j}$ *for all* $j \in \{1,2,...,m\}$.

Suffix, Prefix and Large Suffix of an Episode: The *suffix* of an episode α is defined as an episode composed by the last element of α, the *prefix* of α is the episode composed by all elements in α without the last one and the *large suffix* is the episode composed by the elements in α without the first one. That is, if $\alpha = \langle a_1,a_2,...,a_m\rangle$, then the $suffix(\alpha) = \langle a_m\rangle$, the $prefix(\alpha) = \langle a_1,a_2,...,a_{m-1}\rangle$ and the *large suffix* $lsuffix(\alpha) = \langle a_2,...,a_m\rangle$.

Anti-monotonicity: It is a common principle that frequency measure methods should obey in frequent pattern mining. This principle says that the frequency of an episode must be less or equal to the frequency of its sub-episodes i.e., two episodes α and β in a sequence where $\alpha \subseteq \beta$ follow the principle of anti-monotonicity if $freq(\beta) \leq freq(\alpha)$.

2.1 Serial Occurrences

Given a sequence of events $S = \langle(e_1,t_1),(e_2,t_2),...,(e_n,t_n)\rangle$, a candidate episode $\alpha = \langle a_1,a_2,...,a_m\rangle$ and a maximal gap between events $max\_gap = k$, Algorithm 1 returns the set of maximal non-redundant occurrences, *maxnO*. First, for $m = 1$, the occurrences of the episode are the same minimal occurrences of the event a_1, $maxnO(S,\alpha,k) = mo(a_1)$. Then, for $m > 1$, $maxnO(S,\alpha,k)$ is obtained by properly joining each occurrence of a_1 with the occurrences of $a_2,...,a_m$ located between the corresponding t_1 to $t_1 + (m-1)k$.

For simplicity let each t_i in S take values from $j = 1,2,...,$ and $t_i = j$ means the i-th data element occurs at the j-th tiemestamp. The algorithm has a two phase structure. In the first phase (lines 4-9), a *list* for each occurrence of a_1, $mo(a_1)(i)$, is created containing the occurrences of the other events (a_j) within the constraint k, where $list.a_1 = mo(a_1)(i)$ and $list.a_j = mo(a_j)$ such that $list.a_{j-1}(1) < mo(a_j) \leq list.a_{j-1}(end) + k$ for $j = 2,...,m$.

In the second phase (lines 11-17), the most proper serial occurrence *sO* is selected from the *list*. The most proper occurrence is composed by the most left occurrence of each event found in $list.a_j$ that meets the restrictions of k between events. This is done starting with the first occurrence of the last event from the *list*, that is $list.a_m(1)$ (line 12) and in an iterative procedure the left most occurrence of the other events within k are located (lines 13-17). Each serial occurrence *sO* is added to *maxnO* (line 18) and constitutes the output of the algorithm.

Note that to search the occurrences of an episode (any size), the method requires only the single event occurrences without using their sub-episodes.

Algorithm 1: *serialMethod.*

Input: An event sequence S, a candidate episode $\alpha = \langle a_1, a_2, ..., a_m \rangle$, the maximal gap k, occurrences of the events in α i.e, $mo(a_1), ..., mo(a_m)$.
Output: The maximal non-redundant occurrences of α, $maxnO(S, \alpha, k)$.
Procedure:

Initialise $maxnO(S,\alpha,k) \leftarrow \{\}$
for $i = 1$ to $|mo(a_1)|$ **do**
  //From each $mo(a_1)$ create a *list* of candidate occurrences.
  if $mo(a_1)(i) \notin maxnO(S,\alpha,k)$ **then**
    $list.a_1 \leftarrow mo(a_1)(i)$
    for $j = 2$ to $|\alpha|$ **do**
      $oc \leftarrow mo(a_j)$ such that $list.a_{j-1}(1) < mo(a_j) \leq list.a_{j-1}(end) + k$ and $mo(a_j) \notin maxnO(S,\alpha,k)$
      if $oc \neq \{\}$ **then**
        $list.a_j \leftarrow mo(a_j)(oc)$
    //From *list* select the most proper occurrence.
    if $size(list) = |\alpha|$ **then**
      $sO \leftarrow list.a_m(1)$
      for $j = m-1$ to 1 **do**
        for $kk = 1$ to $|list.a_j|$ **do**
          if $sO(1) - list.a_j(kk) \leq k$ **then**
            $sO \leftarrow [list.a_j(kk)\ \ sO]$
            break
      Add sO to $maxnO(S,\alpha,k)$

Algorithm 2: *parallelMethod.*

Input: An event sequence S, a candidate episode $\alpha = \langle a_1 \cdot a_2 \cdot ... \cdot a_m \rangle$, the maximal gap k, occurrences of the events in α i.e, $mo(a_1), ..., mo(a_m)$.
Output: The maximal non-redundant occurrences of α, $maxnO(S, \alpha, k)$.
Procedure:

1: Initialise $maxnO(S,\alpha,k) \leftarrow \{\}$
2: $vm \leftarrow unique(mo(a_1), ..., mo(a_m))$
3: **for** $i = 1$ to $|vm|$ **do**
4:   **if** $vm(i) \notin maxnO(S,\alpha,k)$ **then**
5:     //Create a list of likely occurrences
6:     $list \leftarrow vm(i)$ to $vm(i + (m-1)k)$ for all $vm.o \notin maxnO(S,\alpha,k)$
7:     //Sort the most probable serial episode
8:     $\alpha_s \leftarrow unique(list.e)$
9:     **for** $j = 1$ to $|\alpha|$ **do**
10:       $Oaux.\alpha_j \leftarrow list.o$ such that $list.e = \alpha_j$
11:     //Find the most properly occurrence
12:     **if** $\alpha \subset \alpha_s$ **then**
13:       $pO \leftarrow serialMethod(list, \alpha_s, Oaux, k)$
14:       **if** $pO = \{\}$ **then**
15:         **for** $j = 2$ to $|\alpha| - 1$ **do**
16:           $\alpha_s \leftarrow reorder(\alpha_s)$
17:           $pO \leftarrow serialMethod(list, \alpha_s, Oaux, k)$
18:           **if** $pO \neq \{\}$ **then**
19:             break
20:       **if** $pO \neq \{\}$ **then**
21:         Add pO to $maxnO(S,\alpha,k)$

2.2 Parallel Occurrences

In a parallel episode there are no constraints about the partial order of the events. The occurrences of a parallel episode include the occurrences of serial episodes composed by the same events, and its frequency is equal or greater than any serial episode composed by the same event types. Methods to measure frequency based on *Fixed window width* and *Non-overlapped occurrences* have been reported in (Mannila et al., 1997) and (Laxman et al., 2004), respectively.

Given a sequence of events S, a candidate parallel episode $\alpha = \langle a_1 \cdot a_2 \cdot ... \cdot a_m \rangle$ and a maximal gap between events $max\_gap = k$, the Algorithm 2 returns the set of maximal non-redundant occurrences $maxnO$. For $m = 1$, the occurrences of the episode are the same minimal occurrences of the event a_1, $maxnO(S, \alpha, k) = mo(a_1)$. For $m > 1$, the occurrences of all events in α are sorted in a structure vm where $vm.o = unique(mo(a_1), ..., mo(a_m))$ contains the occurrences and $vm.e$ contains the corresponding event types, then the set $maxnO(S, \alpha, k)$ is then obtained from it. For each occurrence $vm.o(i)$ the corresponding set of events located between $tm(i)$ to $tm(i + (m-1)k)$ are evaluated to search the more proper occurrence.

The structure of the algorithm is as follows. Each parallel occurrence of an episode is selected in three phases. In the first phase (line 6), for each occurrence in $vm.o(i)$ that has not been considered in previous occurrences, a *list* with the occurrences between $vm.o(i)$ to $vm.o(i + (m-1)k)$ is created.

In the second phase (lines 8-10), the occurrences of each event are saved in an auxiliary list *Oaux*. The most probable serial episode α_s is extracted (line 8) from $list.e$ using the function *unique*. This function selects the first event of each type in α that appears in $list.e$.

Finally, the most proper occurrence pO is extracted (lines 13-21) using the method for serial episodes with α_s, *list*, *Oaux* and k as inputs. Each parallel occurrence pO is added to *maxnO* (line 21) and constitutes the output of the algorithm.

3 SIGNIFICANT EPISODES

Frequent episodes are those that have a number of occurrences greater than a fixed threshold (min_fr), usually predefined by the user; but only some of them are really significant for knowledge discovery purposes. Relevance of frequent episodes have to be evaluated to find representative connections between events describing *patterns*. Confidence of the episode rule can be used for that purpose. In the following paragraph

Table 1: Number of frequent and maximal episodes, and patterns found in the synthetic sequence for several values of *k*.

	k=0.01 s	*k=0.02 s*	*k=0.03 s*	*k=0.04 s*	*k=0.05 s*
Frequent episodes	28	132	389	1149	4354
Maximal episodes	16	35	103	252	902
Patterns:					
$Q_f \Leftrightarrow \{conf \geq 0.8\}$	0	12	32	89	456
$Q_f \Leftrightarrow \{conf_B \geq 0.5\}$	1	16	67	208	824
$Q_f \Leftrightarrow \{coh \geq 0.8 \wedge\}$	6	9	17	24	46
$Q_f \Leftrightarrow \{conf \geq 0.8 \wedge conf_B \geq 0.5\}$	0	5	20	77	413
$Q_f \Leftrightarrow \{coh \geq 0.8 \wedge conf_B \geq 0.5\}$	1	4	6	13	32
$Q_f \Leftrightarrow \{conf \geq 0.8 \wedge coh \geq 0.8\}$	0	4	5	6	14
$Q_f \Leftrightarrow \{conf \geq 0.8 \wedge coh \geq 0.8 \wedge conf_B \geq 0.5\}$	0	3	4	5	11

this evaluation criteria and two new ones (Level of cohesion and the level of backward-confidence) proposed by the authors are defined.

Confidence of an Episode: The confidence of an episode α, $conf(\alpha)$ is the fraction between the frequency of the episode and the frequency of its prefix (Mannila et al., 1997), (Gan and Dai, 2011). The episodes whose confidence is greater than a threshold, min_conf, are called episode rules and can be considered relevant for reasoning tasks. These rules can be interpreted as the probability of occurrence of a new episode once its prefix has occurred.

$$conf(\alpha) = \frac{fr(\alpha)}{fr(prefix(\alpha))} \tag{1}$$

Cohesion of an Episode: The cohesion of an episode α, $coh(\alpha)$ is defined as the fraction between the number of serial and parallel occurrences. This index measures the strength of order relation expressed by the serial episode with respect to other episodes in the sequence containing the same events in different order (parallel episodes).

$$coh(\alpha) = \frac{fr\_serial(\alpha)}{fr\_parallel(\alpha)} \tag{2}$$

Backward-confidence of an Episode: Given that an episode α is the **backward-extention** super episode of its **large suffix** (Zhou et al., 2010), we define the backward-confidence of an episode α, $conf_B(\alpha)$ as the fraction between the frequency of the episode and the frequency of its large suffix. This index measures the probability of occurrence of an episode given the frequency of its large suffix i.e., reveals information about the origin of the episode.

$$conf_B(\alpha) = \frac{fr(\alpha)}{fr(lsuffix(\alpha))} \tag{3}$$

Extraction of Patterns: The significance of frequent episodes can be obtained from their corresponding levels of confidence, cohesion and backward-confidence as a quality factor Q_f defined as:

$$Q_f(\alpha) = f(conf(\alpha), coh(\alpha), conf_B(\alpha)) \tag{4}$$

Restrictions of this quality factor will be set by the user according to discovery goals. The criterion can include one or several indexes combined in different ways. As example, a possible index could be defined by $Q_{f\_min} \Leftrightarrow \{conf \geq min\_conf\}$ or $Q_{f\_min} \Leftrightarrow \{conf \geq min\_conf \wedge coh \geq min\_coh \wedge conf_B \geq min\_conf_B\}$.

Finally, to avoid redundant information, only **maximal frequent episodes** with significant Q_f will be retained as patterns, following the criteria proposed by (Doucet and Ahonen-Myka, 2006).

4 EXPERIMENTAL RESULTS

Frequent episodes and patterns are extracted using the new frequency measure method and the indexes proposed in this paper. Some results are presented and discussed using a synthetic sequence as toy example.

Synthetic sequence was generated by embedding two patterns $\langle L,M,N \rangle$ and $\langle E,F,G,H \rangle$ into a random stream of events using an alphabet of 14 event types. The total sequence time is 5 s and 0.01 s is the average time between events. The detail of the sequence generator can be consulted in (Patnaik, 2006).

Table 1 summarises the main result of the cited sequence using as minimum threshold $min\_fr = 20$ occurrences (the frequency of the less frequent event) and several maximal gaps k between events. As quality factor Q_f for pattern extraction, we have fixed $min\_conf = 0.8$, $min\_coh = 0.8$ and $min\_conf_B = 0.5$ (based on the corresponding values of the embedding patterns). All the frequent episodes are evaluated, using one or several indexes, but only the **maximal episodes** are retained to avoid redundant information.

The first row of Table 1 shows the total number of frequent episodes for different values of k. It is observed that their number increases rapidly as k is relaxed. The corresponding number of maximal episodes is show in the second row of the table. Their number is still significantly high, since the number of embedded patterns is only two.

From row three hereinafter, the number of patterns

Table 2: Patterns extracted using $Q_f \Leftrightarrow \{conf \geq 0.8 \wedge coh \geq 0.8 \wedge conf_B \geq 0.5\}$ as selection criteria.

k=0.02 s	*k=0.03 s*	*k=0.04 s*
$\langle L,M,N \rangle$	$\langle L,M,N \rangle$	$\langle L,M,N \rangle$
$\langle G,H,G,H \rangle$	$\langle G,H,G,H \rangle$	$\langle M,N,M,N \rangle$
$\langle E,F,G,H \rangle$	$\langle F,G,F,G \rangle$	$\langle E,F,G,H \rangle$
	$\langle E,F,G,H \rangle$	$\langle E,F,G,F,G \rangle$
		$\langle G,H,G,H,G,H \rangle$

extracted using one or several of the proposed criteria are shown. For k not equal to 0.01 s the number of patterns using the criteria of *min_coh* is smaller than those extracted using criteria of min_conf or min_conf_B, respectively. With combination of two criteria, the best result (smaller number of patterns) is obtained using min_conf and *min_coh*. However, the combination of the three criteria delivers much better results.

Table 2 shows the patterns extracted with the combination of the three criteria ($conf \wedge coh \wedge conf_B$) for *k=0.02 s* to *k=0.04 s*. The two patterns $\langle L,M,N \rangle$ and $\langle E,F,G,H \rangle$ embedded in the sequence were extracted satisfactorily (except for *k=0.01 s*). As the constrain of maximal gap is relaxed (k increases), other frequent patterns involving mainly the frequent events F, G, and H begins to be significant.

This example shows that the proposed indexes of cohesion (*coh*) and backward-confidence ($conf_B$) may be helpful in the selection of the most significant patterns, improving the results obtained by the simple extraction of maximal episodes or episode rules.

5 CONCLUSIONS

The problem of discovering significance of episodes (patterns) has been analysed and two new indexes called cohesion and backward-confidence of the episodes have been proposed to improve pattern discovery from frequent episodes. A new method to find the maximal number of serial and parallel occurrences has also been presented. Experimental results using a synthetic sequence show that both, the indexes and the algorithms proposed, are useful to search significant patterns in sequences of events.

Set the properties of the method as well as assessing their performance against similar frameworks using real and synthetic data, is part of the work in progress.

ACKNOWLEDGEMENTS

This work was supported by the research project *"Monitorización Inteligente de la Calidad de la Energía Eléctrica"* (DPI2009-07891) funded by the *Ministerio de Ciencia e Innovación* (Spain) and FEDER. Also with the support of the *Comissionat per a Universitats i Recerca del Departament d'Innovacio, Universitats i Empresa* of the *Generalitat de Catalunya* and also the European Social Fund under the FI grant 2012FI_B2 00119.

REFERENCES

Agrawal, R. and Srikant, R. (1994). Fast algorithms for mining association rules. In *Int. Conf. Very Large Data Bases (VLDB'94)*.

Agrawal, R. and Srikant, R. (1995). Mining sequential patterns. In *Int. Conf. Data Engineering (ICDE'95)*, pages 3–14.

Casas-Garriga, G. (2003). Discovering unbounded episodes in sequential data. In Lavrac, N., Gamberger, D., Todorovski, L., and Blockeel, H., editors, *Knowledge Discovery in Databases: PKDD 2003*, volume 2838 of *Lecture Notes in Computer Science*, pages 83–94. Springer Berlin / Heidelberg.

Doucet, A. and Ahonen-Myka, H. (2006). Fast extraction of discontiguous sequences in text: a new approach based on maximal frequent sequences. *Proceedings of IS-LTC*, 2006:186–191.

Gan, M. and Dai, H. (2010). A study on the accuracy of frequency measures and its impact on knowledge discovery in single sequences. In *Data Mining Workshops (ICDMW), 2010 IEEE International Conference on*, pages 859–866.

Gan, M. and Dai, H. (2011). Fast mining of non-derivable episode rules in complex sequences. In Torra, V., Narakawa, Y., Yin, J., and Long, J., editors, *Modeling Decision for Artificial Intelligence*, volume 6820 of *Lecture Notes in Computer Science*, pages 67–78. Springer Berlin / Heidelberg.

Iwanuma, K., Ishihara, R., Takano, Y., and Nabeshima, H. (2005). Extracting frequent subsequences from a single long data sequence a novel anti-monotonic measure and a simple on-line algorithm. In *Data Mining, Fifth IEEE International Conference on*, page 8 pp.

Laxman, S., Sastry, P., and Unnikrishnan, K. (2007). Discovering frequent generalized episodes when events persist for different durations. *Knowledge and Data Engineering, IEEE Transactions on*, 19(9):1188–1201.

Laxman, S., Sastry, P. S., and Unnikrishnan, K. P. (2004). Fast algorithms for frequent episode discovery in event sequences. Technical report, CL-2004-04/MSR, GM R&D Center, Warren.

Mannila, H., Toivonen, H., and Verkamo, A. I. (1997). Discovery of frequent episodes in event sequences. *Data Mining and Knowledge Discovery*, 1(3):259–289.

Patnaik, D. (2006). Application of frequent episode framework in microelectrode array data analysis. Master's thesis, Dept. Electrical Engineering, Indian Institute of Science, Bangalore.

Zhou, W., Liu, H., and Cheng, H. (2010). Mining closed episodes from event sequences efficiently. In *Proceedings of the 14th Pacific-Asia Conference on Knowledge Discovery and Data Mining(1)*, pages 310–318.