RASP: Robust Mining of Frequent Temporal Sequential Patterns under Temporal Variations - Supplementary Document

Algorithm 5: Mining Frequent TSP (Size 2)

Input: (1) W: mapping between (event, relative occurrence time) pairs and their WIDs,
(2) F_1: mapping between events and their WIDs,
(3) m_2: max number of TSPs of size 2,
(4) δ: max time gap between consecutive events,
(5) I: tolerance against temporal variations
Output: F_2: mapping between frequent TSPs of size 2 and (WIDs, frequency) pairs

Lines 1-2 of Algorithm 3 of the main paper
foreach $E_1 \in F_1.keys()$ **do**
  foreach $E_2 \in \mathcal{E}\ \cancel{\setminus\{E_1\}}$ **do**
    for $t \leftarrow 0$ ***to*** $\lfloor \delta/I \rfloor$ **do**
      $\alpha \leftarrow (\langle E_1, E_2 \rangle, \langle tI \rangle)$ `// tI = ΔT̄_t (Sect. 4.3)`
      if $E_1 \neq E_2$ **then**
        $W_\alpha \leftarrow F_1[E_1] \cap (\cup_{w=-I}^{I-1} W[(E_2, \min(tI + w, \delta))])$
      else
        $W_\alpha \leftarrow \emptyset$
        for $w \leftarrow -I$ ***to*** $I - 1$ **do**
          if $\min(tI + w, \delta) \neq 0$ **then**
            $W_\alpha \leftarrow W_\alpha \cup W[(E_2, \min(tI + w, \delta))]$
        $W_\alpha \leftarrow F_1[E_1] \cap W_\alpha$
      Lines 8-15 of Algorithm 3 of the main paper
  return F_2

Algorithm 6: Mining Frequent TSP (size i)

Input: (1) W: mapping between (event, relative occurrence time) pairs and their WIDs,
(2) F_{i-1}: mapping between frequent TSPs of size $(i-1)$ and (WIDs, frequency) pairs,
(3) m_i: max number of TSPs of size i,
(4) δ: max time gap between consecutive events,
(5) L: max time span of a TSP,
(6) I: tolerance against temporal variations
Output: F_i: mapping between frequent TSPs of size i and (WIDs, frequency) pairs

Lines 1-2 of Algorithm 4 of the main paper
foreach $\alpha' \in F_{i-1}.keys()$ **do**
  $(\langle E_1, \cdots, E_{i-1} \rangle, \langle \Delta t_1, \cdots, \Delta t_{i-2} \rangle) \leftarrow \alpha'$
  $(W_{\alpha'}, freq_{\alpha'}) \leftarrow F_{i-1}[\alpha']$
  foreach $E_i \in \mathcal{E}\ \cancel{\setminus\{E_1, \cdots, E_{i-1}\}}$ **do**
    for $t \leftarrow 0$ ***to*** $(\lfloor \min(\delta, L - \sum_{j=1}^{i-2} \Delta t_j)/I \rfloor)$ **do**
      $\Delta t_{i-1} = tI$; $t_i = \sum_{j=1}^{i-1} \Delta t_j$
      $\alpha'' \leftarrow (\langle E_2, \cdots, E_i \rangle, \langle \Delta t_2, \cdots, \Delta t_{i-1} \rangle)$
      if $\alpha'' \in F_{i-1}.keys()$ **then**
        $\alpha \leftarrow (\langle E_1, \cdots, E_i \rangle, \langle \Delta t_1, \cdots, \Delta t_{i-1} \rangle)$
        if $E_i \neq E_j, 1 \leq \forall j \leq i - 1$ **then**
          $W_\alpha \leftarrow W_{\alpha'} \cap (\cup_{w=-I}^{I-1} W[(E_i, \min(t_i + w, \delta))])$
        else
          $W_\alpha \leftarrow \emptyset$
          for $w \leftarrow -I$ ***to*** $I - 1$ **do**
            if $\min(t_i + w, \delta) \neq 0$ **then**
              $W_\alpha \leftarrow W_\alpha \cup W[(E_i, \min(t_i + w, \delta))]$
          $W_\alpha \leftarrow W_{\alpha'} \cap W_\alpha$
        Lines 13-20 of Algorithm 4 of the main paper
  return F_i

A ADDITIONAL EXPERIMENTAL RESULTS

Appendix A is attached to the main paper.

B EXTENSION OF RASP TO ALLOW DUPLICATE EVENTS

In the main paper, we restricted TSPs to distinct events to maintain clarity and readability in our descriptions. Allowing duplicate events could have complicated the already complex explanations of our concepts and algorithms. Fortunately, it is technically feasible to extend our methodology to accommodate duplicate events. Below, we provide a detailed description of this extension, emphasizing the modifications (highlighted in blue) required in our concepts and algorithms to accommodate duplicate events.

Changes Required in Preliminary Concepts (Sect. 3.1) If duplicate events are allowed. a *temporal sequential pattern* (TSP) $\alpha = (\langle E_1, E_2, \cdots, E_l \rangle, \langle \Delta t_1, \Delta t_2, \cdots, \Delta t_{l-1} \rangle)$ of size l is defined as a pair of ordered sequences of (a) (potentially duplicate) events and (b) non-negative *time gaps* between two consecutive events. I.e., $E_i \in \mathcal{E}, 1 \leq \forall i \leq l$, $\cancel{E_i \neq E_j, 1 \leq \forall i \neq j \leq l,}$ and $\Delta t_i \in \mathbb{N}_0, 1 \leq \forall i \leq l - 1$.

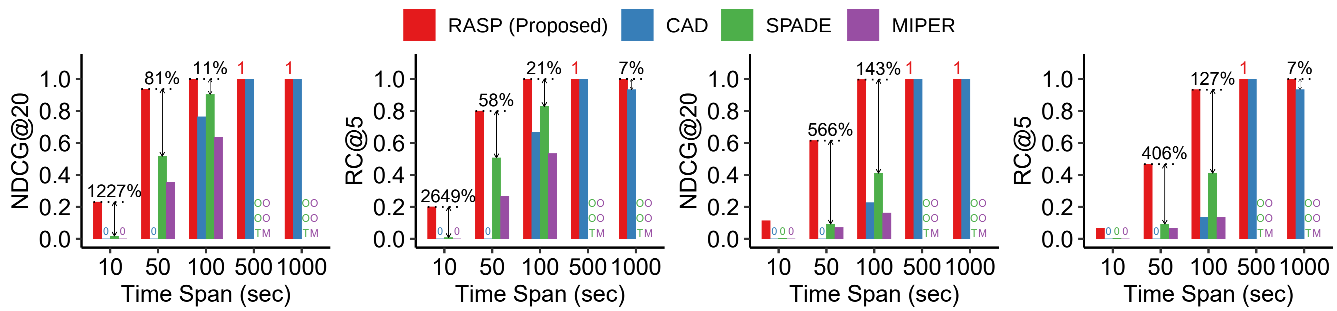
Changes Required in Relaxed TSPs and Duplicated Pattern Matching (Sect. 4.3) Algs. 5 and 6 depict how RASP discovers frequent TSPs of size 2 and larger, respectively, when duplicate events are allowed.

C OTHER EVALUATION METRICS

As additional evaluation metrics, we measure NDCG@20 and RC@5 in each setting and report them in Fig. 17 (**S1. Variation-Free**), Fig. 15 (**S2. Variations**), Fig. 16 (**S3. Event Count**), and Fig. 18 (**S4. Mixed-Easy** and **S5. Mixed-Hard**), respectively. Our proposed method, RASP, demonstrates superior accuracy compared to its competitors across all evaluation metrics and settings.

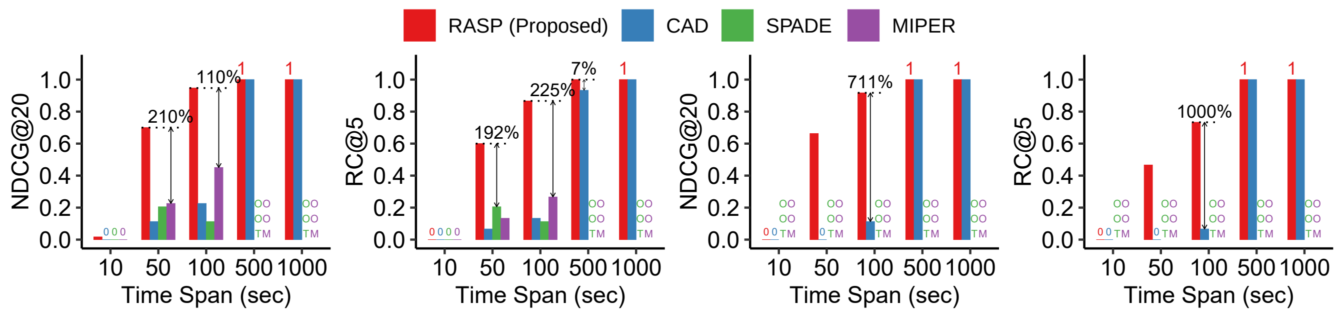
Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn



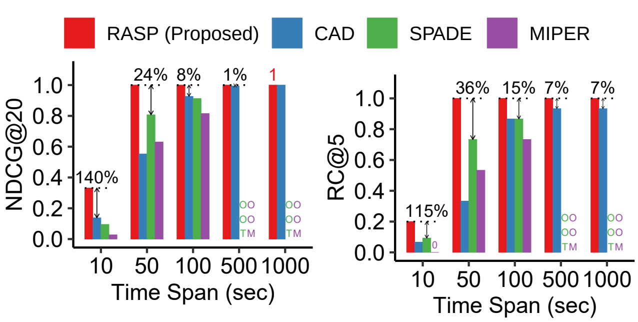
(a) 10 **ms / NDCG@20** **(b)** 10 **ms / RC@5** **(c)** 20 **ms / NDCG@20** **(d)** 20 **ms / RC@5**

Figure 15: S2. Variations. Our proposed method, RASP, exhibits superior accuracy compared to its competitors in the dataset settings with temporal variations with a specified standard deviation.



(a) 100 **events / NDCG@20** **(b)** 100 **events / RC@5** **(c)** 1,000 **events / NDCG@20** **(d)** 1,000 **events / RC@5**

Figure 16: S3. Event Count. Our proposed method, RASP, exhibits superior accuracy compared to its competitors across dataset settings with varying numbers of events.



(a) NDCG@20 **(b) RC@5**

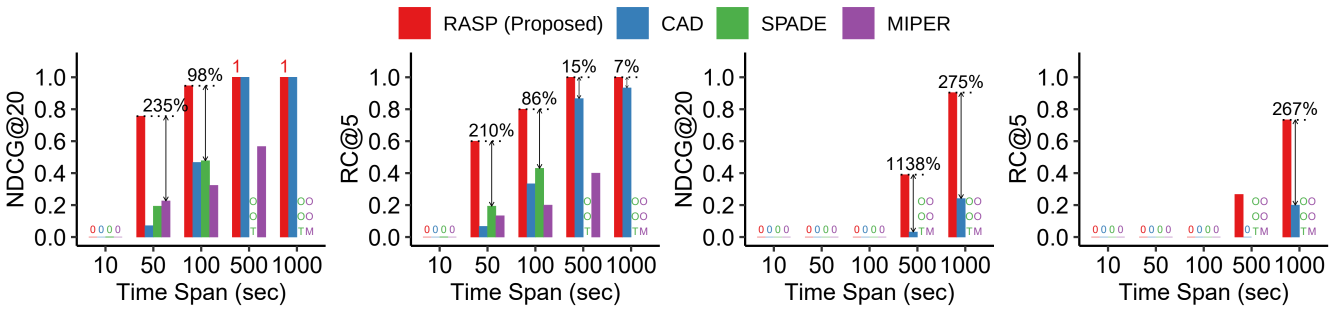
Figure 17: S1. Variation-Free. Our proposed method, RASP, exhibits superior accuracy compared to its competitors in the dataset settings without temporal variations or probabilistic participation.

D VARIOUS BIN SIZES

Regarding the effects of bin sizes on the performance of the considered methods, we present the results for a time span L of 50 seconds in Fig. 19. For the results when L is 100 seconds, refer to Appendix A.3. of the main paper.

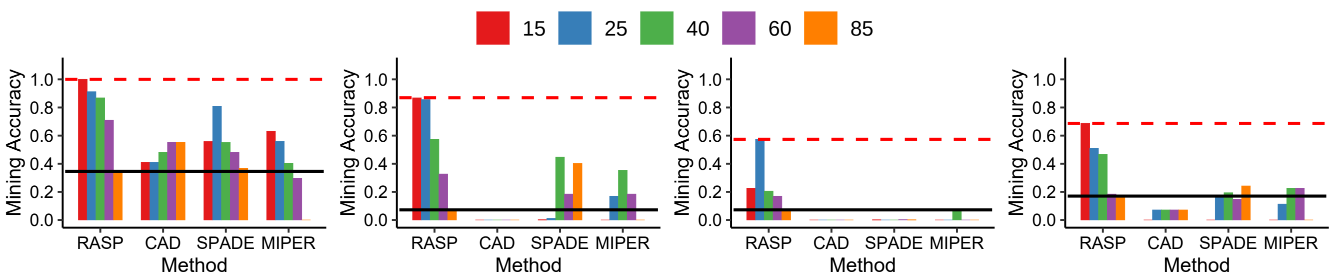
E EXAMPLES OF FREQUENT TSPS IN THE PRECIPITATION DATASET

We examine the two most frequent TSPs identified by RASP from the precipitation dataset (refer to Sect. 5.6.2 for details). As shown in Fig. 20, the AWS locations for the two most frequent TSPs are situated closely near Hallasan on Jeju Island, South Korea.



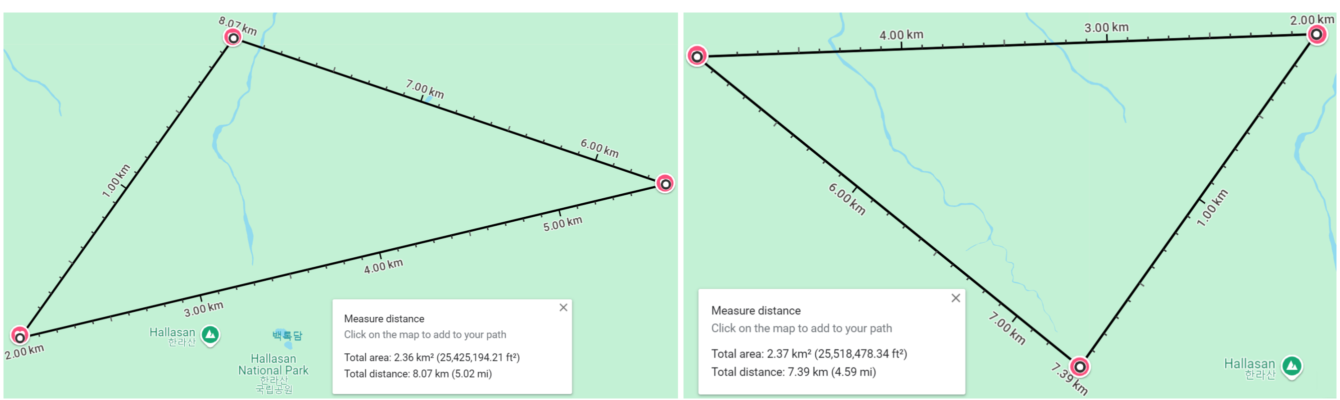
(a) Easy / NDCG@20 **(b) Easy / RC@5** **(c) Hard / NDCG@20** **(d) Hard / RC@5**

Figure 18: S4. Mixed-Easy and S5. Mixed-Hard. Our proposed method, RASP, exhibits superior accuracy compared to its competitors across two dataset settings: S4. Mixed-Easy and S5. Mixed-Hard.



(a) S1. Variation-Free **(b) S2. Variations (10 ms)** **(c) S2. Variations (20 ms)** **(d) S4. Mixed-Easy**

Figure 19: Results with various bin sizes in milliseconds, when the time span L is 50 seconds (see the main paper for the results with a time span L of 100 seconds). Red dotted lines and black solid line show the best and worst performances of RASP. Note that larger time bins tend to benefit CAD, but not necessarily the other methods for which the optimal bin size tends to increase with larger temporal variations. With a proper bin size, RASP performs best in all the settings.



(a) Most Frequent TSP **(b) 2nd Most Frequent TSP**

Figure 20: The AWS locations for the two most frequent TSPs identified by RASP from the precipitation dataset.