Cascading Spatio-Temporal Pattern Discovery

Paper summarized by G3 (Akash Agrawal, Atanu Roy)

Peer-reviewed by G10 (Anuj Karpatne, Vijay Mohan)

Introduction to the paper:

The paper introduces the concept of mining cascading spatio-temporal patterns (CSTPs) which can be defined as partially ordered subsets of event-types which are located together in space and occur serially in time. The authors provide an extension of their previous work on mining CSTPs using directed acyclic graph (DAG) models, where they had developed an interest measure for assessing DAGs and a nested loop algorithm for mining such patterns. Drawing on the motivation from the challenges faced in their previous approach, the paper extends the existing framework by introducing a new pattern mining algorithm based on spatio-temporal partitioning and by performing extensive computational cost analyses of various sub-procedures performed in the overall CSTP mining algorithm. Deductive analysis and empirical evaluations on real world and synthetically generated datasets perform both a quantitative analysis of the performance gain of the algorithm and a qualitative assessment of the insights obtained about the real-world problem by analyzing their results.

Review of the presentation on the paper:

The presentation slides and the narrative provide a coherent summary of the problem in a brief and concise fashion. In their thorough reading of this paper, they came across some minor printing mistakes in one of the figures in the paper that were conceptually non-coherent with the text, and which were communicated to the authors. Also, they suggested using network distance as an improvement and performing more exhaustive experiments to prove the efficacy of the proposed model and to provide a comparative performance analysis of the different steps involved in the algorithm.

However, the presentation lacks on addressing to certain key points of the paper which would have given a better insight into the paper. Neglecting the minor spelling and editing mistakes, the following issues need to be re-looked at and improved:

- The paper extends on an existing work of the authors by developing a more detailed perspective on the efficacy of the previous algorithm and addressing to some of the challenges they had faced in their previous approach. However, the presentation lacks in bringing out this optimization aspect of their work. Their section on the Problem Statement dwells more on introducing the problem of the previous approach without any reference to the improvement being addressed to in the paper.
- The challenges of the previous approach that motivate the development of the new algorithm have been skipped from the slides maybe due to time constraints. The narrative also lacks in explaining the motivations drawn for developing the new algorithm as different from the previous approach.
- For the Contributions Section, it is more relevant to state the contributions of the current work
only for providing a more clear and less ambiguous description of the claims of the paper. The contributions of the previous work can be included in the section on Proposed Approach, where it can help in explaining the new approach in a better fashion.

- The motivation for using filters (for addressing Challenge1: reducing the exponential cardinality of patterns) has not been explicitly stated, which could have helped in developing a connected story in the summary.

- The motivation behind using better evaluation procedures of the interest measure of a pattern (for addressing to the empirically observed phenomena that interest measure computation is significantly more expensive than candidate set generation) has not been mentioned separately.

- The key focus of the paper is on developing better interest measure evaluation techniques of a pattern and addressing to this issue, they had introduced a partitioning based CPI evaluation. However, both the summary and the slides lacks in mentioning any reference to the new method (TKDE CSTPM).

- The statistical significance of CPI draws from its upper-boundness to the space-time K-Function, which is a ST statistical measure. Since one of the key challenges of their previous approach (Challenge3) was to make a balance between computational complexity and statistical significance, it is relevant to include this point in the presentation, which is not the case currently. Anti-monocity, and its relevance to this problem can also be explained briefly.

- Novelty hasn't been addressed in both the slides and the narrative as they haven't provided any reference to the related work in this area in the recent literature. A comparative analysis should be performed between the proposed model and the existing methods (for eg: Bayesian networks, association analysis, maximal independent set computation, moving object tracking etc.) which should present the advantages of the new model over other models, highlighting the novelty of the work.

- Since the paper presents an extension of the previous approach, it would be worthwhile to provide an example explaining the problem at hand and the challenges that are being addressed in the new approach. This can provide the necessary motivation for introducing the new pattern mining algorithm.

- The algebraic costs of the two methods used for CPI computation and the computational advantage of using filtering as opposed to no-filtering appear to be prime focus of the paper and should be mentioned in the summary.

- An insight into the real-world example results obtained by the proposed approach would be very useful to present as it would motivate its future applications and extensions.

- In the presentation, the previous work can be clubbed into 1 or 2 slides, focusing more on the optimization of the existing problem being addressed to in the paper. Similarly, in the narrative the definitions can be reduced by providing an example in the Proposed Approach Section, giving more focus on the current work.