Poster ID: 25 | Discovering Periodic-Frequent Patterns in Uncertain Temporal Databases

PALLA Likhitha, Supervisor: RAGE Uday Kiran

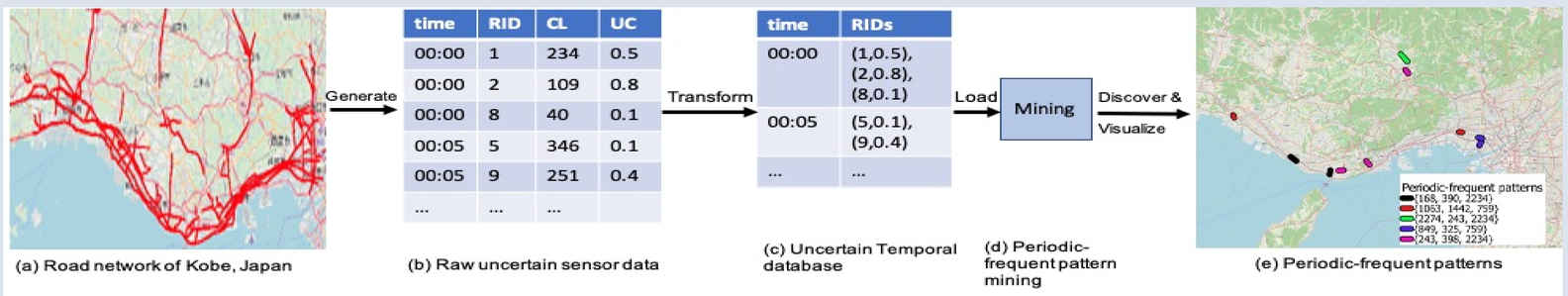
Database Systems Lab, University of Aizu

Abstract

- Uncertain data is naturally by real-world applications, such as sensor networks and satellite imagery data.
- Useful information that can facilitate the end-users of Society 5.0 to achieve socio-economic development lies hidden in this data.
- We proposed a novel model and an efficient algorithm to discover user interest-based patterns, called periodic-frequent patterns, in uncertain temporal databases.

Introduction

- The data generated by real world applications naturally exist as temporal database with uncertainity.
- Useful information is hidden in this data.
- Periodic-frequent pattern mining aims to discover periodically occurring hidden patterns in the data.



Definition and Challenges

Problem definition: Given an uncertain temporal database, the objective of our model is to find all patterns that have *periodicity* no more than *maximum periodicity* (*maxPer*) and *support* no less than *minimum support* (*minSup*).

Challenges:

1. **Lack of mathematical model.** Existing studies disregarded the temporal occurrence information of the items in the data.
2. **Huge search space.** The search space of periodic-frequent pattern mining is $2^n - 1$, where n represents the total number of items. Searching this huge search space is a challenging task.
3. **Need for algorithm.** Previous algorithms do not take into account the uncertain nature of data.

Experimental Results

Since there exists no algorithm to find periodic-frequent patterns in an uncertain temporal database, we compare UPFP-growth against a naïve algorithm and show that UPFP-growth is efficient. The naïve algorithm involves the following two steps: (*i*) finding all frequent patterns in an uncertain temporal database using PUF algorithm and (*ii*) generating periodic-frequent patterns from frequent patterns by performing another scan on the database.

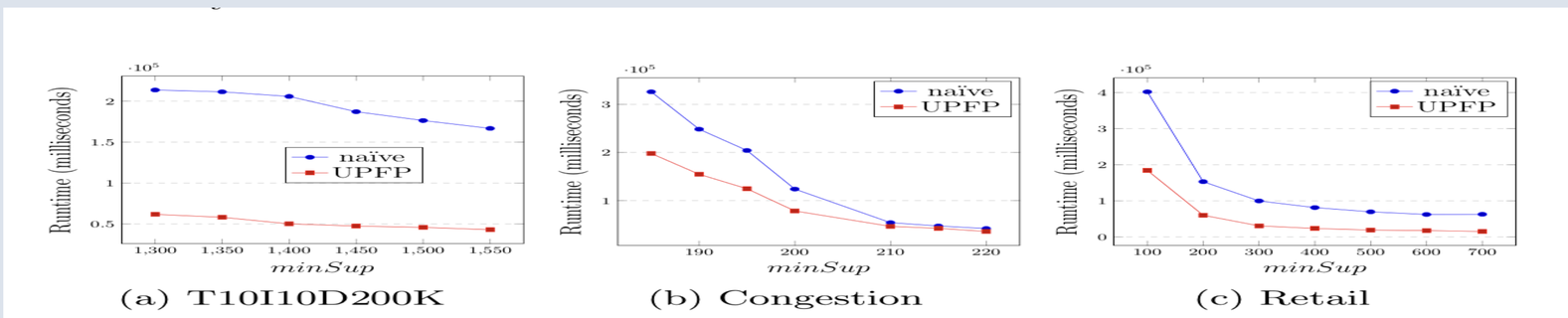


Fig. 1. Runtime requirements of naïve and UPFP-growth algorithms

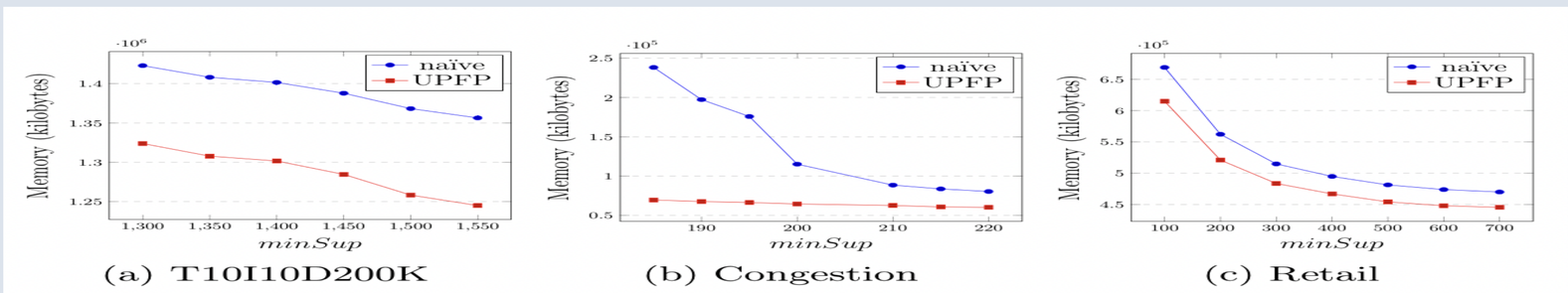
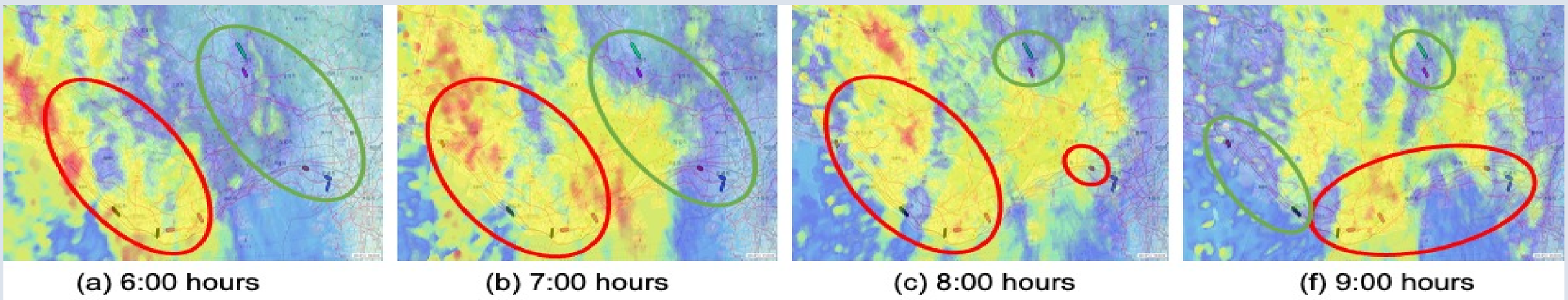


Fig. 2. Memory requirements of naïve and UPFP-growth algorithms

Case Study

Below figure shows the application in which hourly rainfall data of typhoon Nangka was interpolated on the generated patterns. Road segments in red circles require attention, while road segments in green circles require considerably less attention from users. Such information can be helpful to the traffic control room users for diverting the traffic and suggesting patrol routes to reduce accidents. Here we have demonstrated the usefulness of our model using historical data. However, similar studies can be conducted using predicted congestion and rainfall data.



Conclusion

- Proposed a novel model in [1] to find periodic-frequent patterns in an uncertain temporal database. Two constraints minimum support, maximum periodicity were utilized to determine a patterns interestingness in the database.
- Introduced to two tighter-upper bound measures to effectively reduce the search space and computational cost of finding desired patterns.
- Introduced new data structure (UPFP-tree) and a pattern growth algorithm(UPFP-growth) to find all the periodic-frequent patterns.
- Experimental results demonstrate that our algorithm is memory and runtime efficient.

References

[1] R. Uday Kiran, P. Likhitha, M.-S. Dao, K. Zettsu, and J. Zhang, "Discovering periodic-frequent patterns in uncertain temporal databases," in *Neural Information Processing*, T. Mantoro, M. Lee, M. A. Ayu, K. W. Wong, and A. N. Hidayanto, Eds., Cham: Springer International Publishing, 2021, pp. 710–718, ISBN: 978-3-030-92307-5.