Chapter VIII

Mining Flow Patterns in Spatio-Temporal Data

Besides topological patterns, another class of useful spatio-temporal patterns is spatial sequence patterns. In Chapter VI, we introduce the topological patterns to find the intra-relationships of events in a time window. These events may relate to each other in a star, clique, or star-clique manner within the given time window. However, the patterns cannot disclose the inter-relationships of events in different time windows. For example,

“Forest fire always occurs at region $R_1$ prior to the occurrence of haze in nearby region $R_2$.” or

“Forest fire always occurs at a region prior to the occurrence of haze in its Northeastern nearby regions.”

Here, we link the event “fire at $R_1$” to the event “haze in $R_2$”. Such information cannot be obtained by spatial patterns, temporal patterns, or topologi-
cal patterns alone. In other words, the focus of spatial sequence patterns is to describe how the observation of one event in some location implies the occurrence of another event in a second location, or how changes of events in one location can affect the events in another location. Such insights are critical for decision markers.

Further investigation reveals that this “flow” of events can be described in terms of the absolute location coordinates (region $R_2$), or the relative location coordinates (Northeastern regions). We call the absolute location patterns the flow pattern; and the relative location patterns the generalized spatio-temporal patterns.

The discovery of spatial sequence patterns is challenging because of the potentially large search space and the large number of candidates. Naïve incorporation of spatial information into existing sequence mining algorithms does not work. This calls for new spatial sequence pattern mining algorithms.

In this chapter, we describe flow patterns and the design of the algorithm called FlowMiner to find such flow patterns. FlowMiner incorporates a new candidate generation algorithm and employs various optimization techniques for better efficiency. The discovery of generalized spatio-temporal patterns will be described in the next chapter.

The rest of the chapter is organized as follows. We first give the notations and terminologies used, and define the problem of mining flow patterns. Then we illustrate the concept of flow patterns and the design of a new candidate generation algorithm which generates candidates by utilizing only length-2 sequences, instead of length $k-1$ sequences. By combining this with a depth-first search strategy and taking into account some spatial constraints, we are able to eliminate the need to generate large number of candidates. We introduce a disk-based algorithm called FlowMiner that incorporates the new candidate generation algorithm and employs various optimization techniques to improve its efficiency. Performance study shows that FlowMiner is both scalable and efficient, and experiments on real-life datasets also reveal some interesting flow patterns.

Notations and Terminologies

Spatio-temporal databases capture both the time and space dimensions. First, we divide time into disjoint time windows of length $W$. Each time window
Discovering Personalized Novel Knowledge from Text
Yi-fang Brook Wu and Xin Chen (2009). *Handbook of Research on Text and Web Mining Technologies* (pp. 301-313).
[www.igi-global.com/chapter/discovering-personalized-novel-knowledge-text/21731?camid=4v1a](http://www.igi-global.com/chapter/discovering-personalized-novel-knowledge-text/21731?camid=4v1a)