学 位 論 文 内 容 の 要 旨

博士の専攻分野の名称　博士（情報科学）　氏名　耿　暁亮

学 位 論 文 題 名

Studies on Depth-First Mining Algorithms for Maximal Flock Patterns from Large Trajectory Data

（大規模軌跡データからの極大群パターン発見のための深さ優先アルゴリズムに関する研究）

As the period of big data comes, researchers focus their eyes on novel types of data, such as streaming data and spatio-temporal data. GPS trajectory data as one of these kinds of novel data becomes more attractive than before. Since mining of trajectory data has different characteristics from traditional transaction data mining research, trajectory mining has attracted a great deal of attention for recent years. Flock pattern mining is one of the trajectory data mining topic introduced by Gudmundsson et al. in 2006. Flock patterns are a class of spatio-temporal patterns that represent a groups of moving objects which are close to each other in a given time segment. They are useful in detecting a group of highly correlated entities combining spatio-temporal features.

In flock pattern mining, however, a huge number of redundant patterns with shorter lengths and fewer occurrences can be generated, which may degenerate the efficiency and comprehensiveness. One of the possible approaches to this problem is to find only a subset of reprsentative patterns rather than all patterns. To overcome these difficulties, in this thesis, we introduce the classes of length-maximal and size-maximal flock patterns. Then, we devise efficient mining algorithms that find these classes of flock patterns from large trajectory data. Particularly, we present a variety of depth-first mining algorithms based on pattern-growth approach for finding maximal flock patterns, which are efficient in terms of both time and space.

In Chapter 3, we focus on mining of rightward length-maximal flock patterns. A rightward length-maximal flock pattern (RFP, for short) is such a maixmal flock pattern that cannot be extended rightward further without changing their occurrences. After introducing a basic

algorithm FPM (Flock Pattern Miner), we present a depth-first algorithm RFPM (Rightward-maximal Flock Pattern Miner) for finding all RFPM patterns with given radius, and then a faster algorithm G-RFPM that uses a geometric index for data points to speed-up the search for RFPs.

In Chapter 4, we extended the results in Chapter 3 for the class of two-sided length-maximal flock patterns. Introduction of RFPs eliminates some redundant patterns with the same start points, while it may still include other shorter patterns with the different start points. To exclude these redundancy, we introduce the class of twosided length-maximal flock patterns (MFP, for short), which is a maixmal flock pattern that cannot be extended either leftward or rightward. Consequently, we see that the number of MFPs is no more than the number of RFPs in any trajectory data. For mining MFPs, we first present the algorithm FPM-M (FPM for Maximal Patterns) based on rejection of redundant patterns. Next, we present faster algorithms FPM-MP (FPM-M with Partition), which is based on recursive partition of patterns, and FPM-MG (FPM-M with Geometric Index) with a spatial index. All algorithms find all length-maximal flock patterns with given radius appearing in trajectory data.

In Chapter 5, we study size-maximal flock pattern mining problem. We present an algorithm DFM (Depth-First Miner) for mining a subclass of size-maximal envelope patterns, called size-maximal and rightward length-maximal flock patterns (RSFP, for short), under given constraints from an input trajectory database. In reality, the algorithm works on equivalent representation of size-maximal and rightward length-maximal flock patterns, called rightward length-maximal envelope patterns (REVPs, for short). The algorithm combines depth-first search, efficient closure computation, and hash-based duplication check to achieve complete mining of all RSFPs.

Finally, we conclude this thesis and discuss future researches in Chapter 6. In summary, we studied the problem of mining various notions of maximal flock patterns, and presented efficient depth-first mining algorithm that find all maximal patterns in an input trajectory database for these classes of patterns. In experiments on artificial and real trajectory data, our algorithms were correctly and efficiently find all flock patterns.