

ETSDM011- 2017 Ant Colony System Sanitization Approach to Hiding Sensitive Itemsets

Abstract

In recent years, privacy-preserving data mining (PPDM) has received a lot of attention in the field of data mining research. While some sensitive information in databases cannot be revealed, PPDM can discover additional important knowledge and still hide critical information. There are different ways to approach this exhibited in previous research, which applied addition and deletion operations to adjust an original database in order to hide sensitive information. However, it is an NP-hard problem to find an appropriate set of transactions/itemsets for hiding sensitive information. In the past, evolutionary algorithms (EAs) were developed to hide sensitive itemsets by building an appropriate database. Genetic-based algorithms (GAs) and a particle swarm optimization (PSO)-based algorithm, proposed in previous works, not only hide sensitive itemsets but minimize the side effects of sanitization processes. In this paper, an ant colony system (ACS)-based algorithm called ACS2DT is proposed to decrease side effects and enhance the performance of the sanitization process. Each ant in the population will build a tour for each iteration, and each tour indicates the deleted transactions in the original database. The proposed algorithm introduces a useful heuristic function to conduct each ant to select a suitable edge (transaction) for the current situation and also designs several termination conditions to stop the sanitization processes. The proposed heuristic function applies the pre-large concept to monitor side effects and calculates the degree of hiding information to adjust the selecting policy for deleted transactions. The experimental results show that the proposed ACS2DT algorithm performs better than the Greedy algorithm and other two evolutionary algorithms in terms of runtime, fail to be hidden (F-T-H), not to be hidden (N-T-H), not to be generated (N-T-G) and database similarity (DS) on both real-world and synthetic datasets.



