有效率的循序樣本探勘系統及其在樂透開獎預測之應用

To speed up sequential pattern mining tasks for large scale database, in this research, several improvements on the mining bottleneck are proposed and an Efficient Sequential Pattern Mining (ESPM) are then presented based on these improvements. The features of ESPM are described as follow. To reduce the number of query and matching in the database, a modified vertical data format is used to construct an image of the database in memory. Besides, to prevent the mining process from the candidate itemset explosion, the hash concept is adopted for quickly determine whether an itemset is frequent or not. In the simulation experiments, ESPM are tested for the mining in a database containing up to 900,000 randomly-generated transactions. The results show the effectiveness of ESPM. In addition, ESPM is also applied to the prediction of lottery numbers. The mining result shows that AnB templates are not useful to the lottery prediction.