Privacy Preserving Scheme for Sequential Pattern Mining Over Fragmented Progressive Databases

Durga Toshniwal*, Amruta Mhatre**, Keshavamurthy B N***

Frequent Sequential Pattern Mining, is a data mining technique used to find interesting patterns in a large collection of data items. The databases used for mining patterns in real life applications are progressive databases, which are a generalized model providing dynamic addition and deletion of data for efficient mining operations.

Sometimes a group of local market players may be interested in mining trends by pooling in their individual data. However, the shared data may disclose some information which might be against the privacy policies of these collaborating parties or may be of strategic importance for some party. The proposed work proposes a set of algorithms for finding sequential patterns from distributed databases while preserving privacy. The work aims at maintaining the privacy of the data and patterns mined with minimal effect on accuracy of the results.

In this work, the algorithms address the two main types of fragmentation (viz. Vertical, Horizontal). In this work, we use cryptographic and randomization techniques to achieve privacy preservation. The work also proposes an idea to suppress sensitive sequential pattern mining results which is a novel sanitizing technique in progressive databases.

Field of Research: Data Mining, Privacy preservation, progressive databases, sequential pattern mining.

* Durga Toshniwal, Assistant Professor, Department of Electronics and Computer Science, Indian Institute of Technology, Roorkee, Uttarakhand, India. Email: durgafec@iitr.ernet.in
** Amruta Mhatre, Department of Electronics and Computer Science, Indian Institute of Technology, Roorkee, Uttarakhand, India. Email: amruta.01@gmail.com
*** Keshavamurthy B N is a Research Scholar at D Department of Electronics and Computer Science, Indian Institute of Technology, Roorkee, Uttarakhand, India.