SMART REACH
(S/W CREATORS & TRAINERS)

CERTIFIED ISO 9001:2008 COMPANY

Ph: 9585554590, 9585554599
Email: support@salemsmartreach.com
URL: www.salemsmartreach.com

Mining Probabilistically Frequent Sequential Patterns in Large Uncertain Databases

Abstract:

Data uncertainty is inherent in many real-world applications such as environmental surveillance and mobile tracking. Mining sequential patterns from inaccurate data, such as those data arising from sensor readings and GPS trajectories, is important for discovering hidden knowledge in such applications. In this paper, we propose to measure pattern frequentness based on the possible world semantics. We establish two uncertain sequence data models abstracted from many real-life applications involving uncertain sequence data, and formulate the problem of mining probabilistically frequent sequential patterns (or p-FSPs) from data that conform to our models. However, the number of possible worlds is extremely large, which makes the mining prohibitively expensive. Inspired by the famous PrefixSpan algorithm, we develop two new algorithms, collectively called U-PrefixSpan, for p-FSP mining. U-PrefixSpan effectively avoids the problem of "possible worlds explosion", and when combined with our four pruning and validating methods, achieves even better performance. We also propose a fast validating method to further speed up our U-PrefixSpan algorithm. The efficiency and effectiveness of U-PrefixSpan are verified through extensive experiments on both real and synthetic datasets.