

Abstract

Khaled Mahar

NewSpirit: Sequential pattern mining with regular expression constraints

the significant growth of sequence database sizes in recent years increase the importance of developing new techniques for data organizationquery processing. discovering sequential patterns is an important problem in data mining with a host of application domains including medicine, telecommunications,the world wide web. conventional mining systems provide users with only a very restricted mechanism (based on minimum support) for specifying patterns of interest. for effectivenessefficiency consideration, constraints are essential for many sequential applications. in this paper, we give a brief review of different sequential pattern mining algorithms,then introduce a new algorithm (termed newspirit) for mining frequent sequential patterns that satisfy user specified regular expression constraints. the general idea of our algorithm is to use finite state automata to represent the regular expression constraintsbuild a tree that represents all sequences of data which satisfy these constraints by scanning the database of sequences only once.