

CLUSTERING OF WEB USAGE DATA USING FUZZY TOLERANCE ROUGH SET SIMILARITY AND TABLE FILLING ALGORITHM

T. VIJAYA KUMAR & H. S. GURUPRASAD

Department of IS & E, BMS College of Engineering, Bull Temple Road, Bangalore, Karnataka, India

ABSTRACT

Web Usage Mining is the application of data mining techniques to learn usage patterns from Web server log file in order to understand and better serve the requirements of web based applications. Web Usage Mining includes three most important steps namely Data Preprocessing, Pattern discovery and Analysis of the discovered patterns. One of the most important tasks in Web usage mining is to find groups of users exhibiting similar browsing patterns. Grouping web transactions into clusters is important in order to understand user's navigational behavior. Different types of clustering algorithms such as partition based, distance based, density based, grid based, hierarchical and fuzzy clustering algorithms are used to find clusters from Web usage data. In this paper we propose an approach for clustering Web usage data based on Fuzzy tolerance rough set theory and table filling algorithm. First, we have constructed the sessions using concept hierarchy and link information. The similarity between two sessions is approximated by using Rough set tolerance relation. The tolerance relation is reformulated into equivalence relation using fuzzy tolerance. Then the clusters are obtained by using modified table filling algorithm. We provide experimental results of Fuzzy rough set similarity and table filling algorithm on MSNBC web navigation data set. In this paper, we have considered the server log files of the Website www.enggresources.com for overall study and analysis.

KEYWORDS: Web Usage Mining, Concept Hierarchy, Website Ontology, Rough Set Similarity, Fuzzy Tolerance, Table Filling Algorithm