ABSTRACT:

The web user sessions are clustered with incorporating the sequence of web page visits. A sequence-based clustering is developed by proposing new sequence representations and new similarity measures. The resulting sequence representation allows for calculation of similarity between web user sessions and then, can be used as input of clustering algorithms. This study proposed a hybrid prediction model (HyMFM) that integrates Markov model, Association rules and Fuzzy Adaptive Resonance Theory (Fuzzy ART) clustering together. The three approaches are integrated to maximize their strengths. A series of experiments was conducted to investigate whether, clustering performance is affected by different sequence representations and different similarity measures. This model could provide better prediction than using each approach individually.