

Text fragment extraction using incremental evolving fuzzy grammar fragments learner

ABSTRACT

Additional structure within free texts can be utilized to assist in identification of matching items and can benefit many intelligent text pattern recognition applications. This paper presents an incremental evolving fuzzy grammar (IEFG) method that focuses on the learning of underlying text fragment patterns and provides an efficient fuzzy grammar representation that exploits both syntactic and semantic properties. This notion is quantified via (i) fuzzy membership which measures the degree of membership for a text fragment in a semantic grammar class and (ii) fuzzy grammar similarity which estimates the similarity between two grammars (iii) grammar combination which combines and generalizes the grammar at a minimal generalization. Terrorism incidents data from the United States World Incidents Tracking System (WITS) are used in experiments and presented throughout the paper. A comparison with regular expression methods is made in identification of text fragments representing times. The application of text fragment extraction using IEFG is demonstrated in event type, victim type, dead count and wounded count detection with WITS XML-tagged data used as golden standard. Results have shown the efficiency and practicality of IEFG.

Keyword: Evolving system; Extensible markup language (XML); Fragment extraction; Fuzzy sets; Grammar fragments; Incremental learning; Instance matching; Tagging; Text mining