**ABSTRACT**

Data mining is an essential step in the process of knowledge discovery in databases. With a large number of patterns generated by using data mining approaches, how to effectively use and update these patterns is still an open research issue. Text mining is the discovery of interesting knowledge in text documents. It is a challenging issue to find accurate knowledge (or features) in text documents to help users to find what they want. Since most existing text mining methods adopted term-based approaches, they all suffer from the problems of polysemy and synonymy. More text sequences are being generated in various forms. To discover valuable knowledge from a text sequence, the first step is usually to extract topics from the sequence with both semantic and temporal information, which are described by two distributions, respectively: a word distribution describing the semantics of the topic and a time distribution describing the topic’s intensity over time. Plagiarism with Text mining attempts to discover new, previously unknown information by applying techniques from natural language processing and data mining. Our system find exact plagiarism count for different documents.