

ETSBD005 - β Algorithm: A New Probabilistic Process Learning Approach for Big Data in Healthcare

Abstract

In this paper, a new process learning framework that is based on probabilistic learning and predicate logic is proposed. The input of this framework is a set of log files, and the output is a probabilistic predicate-based workflow that describes the process. This paper targets a methodology of learning processes given data and the learning algorithm finds out the logical operators that bind the events described in data and model it using predicate logic. While building the process, the probability of every event and the probabilities of the relationship between events are calculated. The learning process is an ongoing process, which means after learning when feeding the system with a new set of log files, the algorithm takes the previously learned process and it is set of probabilities as a starting state and starts modifying them based on the newly learned log files. This feature is very essential for those applications that integrate and interact with bigdata since for bigdata, starting the learning process from the beginning for every new set of data is not feasible. In this paper, the assumption is that log files are event-based, and every event is associated with its time of occurrence. Any event could have multiple occurrence times throughout the log files. The framework provides an optimal general definition of a process that is described by those log files. The process could change schematically or with respect to behavior when learning a new set of logs. In order to achieve what is described, a dependency matrix needs to be learned, and then the probability matrix is calculated. The outcome of the two matrices is a predicate-based workflow. Workflows can easily be described by Petri nets and Petri nets can map to predicate logic. The reason to convert the workflow into the knowledge base is the ability to infer new facts from given facts we conclude from log files. In this paper, we integrate a modification to α algorithm with the framework in order to describe dependencies and probability of occurrences of events.

Index Terms Big data, α algorithm, event logs, process mining, healthcare.



Maruthi Plaza 91/6, TC Palya Main road,
Next to RK Apartments, Ramamoorthy Nagar,
Bangalore-560025.



9543218650



ieeeprojects@eminents.in