Looking into the Possibility for Designing Normal Distribution Based Dissimilarity Measure to Discover Time Profiled Association Patterns

Vangipuram Radhakrishna
Department of Information Technology
VNR VJIET, Hyderabad
India
radhakrishna_v@vnrvjiet.in

Shadi. A. Aljawarneh
Software Engineering Dept
JUST
Jordan
saaljawarneh@just.edu.jo

Vinjamuri Janaki
Professor, Computer Science and Engineering
Vaagdevi College of Engineering
Warangal, India
janakicse@yahoo.com

P.V.Kumar
Professor, Computer Science and Engineering
UCOE, Osmania University (Retd.)
Acharya Institute of Technology, Bangalore
pvkumar58@gmail.com

Abstract—This research addresses the design of a novel dissimilarity measure for mining similar patterns from time stamped temporal databases applying the concept of standard score and normal distribution. The basic idea behind the design of dissimilarity measure is to use and transform supports to z-space and compute the probability of z-score of temporal patterns. The probability is obtained using normal distribution chart. The objective has been to design a normal distribution based dissimilarity measure which can be used to discover all valid similarity- profiled temporal association patterns.

Keywords—normal distribution; time stamp; standard score; seasonal pattern; association; temporal

I. INTRODUCTION

The design of dissimilarity measures proposed in previous works is not based on the normal distribution concept. In this paper, dissimilarity measure which is based on the concept of normal distribution concept is introduced. i.e the design of dissimilarity measure is now extended to suit the possibility of mining time profiled temporal association patterns through applying the normal distribution concept. The basic idea is to transform the patterns support value and their support sequences into their equivalent z-score value and z-score sequences. For these z-score sequences, the probability value is computed using normal distribution chart. Finally, temporal patterns are expressed as sequences of z-score probability values.

Dissimilarity measure proposed in this paper is thus based on the normal distribution concept and here we study the possibility of extending distance measures proposed in our previous works [12, 13], but now in the context of normal distribution.

Section II outlines some of the previous works in the literature. Sections III and IV discuss the design and analysis of normal distribution based dissimilarity measure proposed for mining time profiled temporal patterns. Section-V outlines normal distribution. Section VI gives naïve algorithm and the last section-VII concludes this paper.

II. RELATED WORKS AND PRELIMINARIES

Similarity profiled temporal association pattern mining is one of the topics of wide research interest in the context of temporal data mining. The pioneering work to address the solution in this direction is by the authors, Jin Soung Yoo and Shashi Shekhar [1-3] using Euclidean distance measure. This research is inspired from [1-3, 19, 20]. From the extensive literature survey performed and also to the best of our knowledge, there are no significant findings recorded in the literature in the direction of proposing new measures to address the above said problem. This fact has actually motivated us to come up with new similarity measures, so that these measures can be used to retrieve all valid similar temporal patterns w.r.t any chosen reference pattern. Some of our earlier works [4-10] propose new similarity (or dissimilarity) measures and approaches to estimate the supports of temporal association patterns. This work extends our previous research by proposing a new dissimilarity measure for retrieving all possible and valid time profiled temporal association patterns from the given input time stamped temporal dataset.

In [4], we come up with a dissimilarity measure for mining temporal association patterns, all those patterns whose