Advances in Intelligent Systems and Computing 828

Anand J. Kulkarni
Suresh Chandra Satapathy · Tai Kang
Ali Husseinzadeh Kashan *Editors*

Proceedings of the 2nd International Conference on Data Engineering and Communication Technology

ICDECT 2017

Springer

Data Engineering and Communication Technology

Ali Husseinzadeh Kashan

Series Title
Advances in Intelligent Systems and Computing

DOI
https://doi.org/10.1007/978-981-13-1610-4

Publisher
Springer Singapore

eBook Packages
Engineering, Engineering (R0)

Copyright Information
Springer Nature Singapore Pte Ltd. 2019

Softcover ISBN
978-981-13-1609-8
Published: 04 October 2018

eBook ISBN
978-981-13-1610-4
Published: 03 October 2018

Series ISSN
2194-5357

Series E-ISSN
2194-5365

Frequent Itemsets in Data Streams Using Dynamically Generated Minimum Support

Shankar B. Naik and Jyoti D. Pawar

Abstract We have presented an approach that generates frequent itemsets from data stream. The itemsets are compressed and then stored in the memory. The decision to whether or not compress an itemset is based on the utility of the itemset. In this chapter, the utility of an itemset is defined in terms of the amount of memory saved by its compression. Beside this, we have presented an approach to dynamically generate the value of minimum support threshold based on the data in the data streams. It avoids having a fixed minimum support threshold throughout the data stream. Since the value is generated from the latest elements in the data stream, it suits to be an appropriate measure to separate the frequent itemsets from the non-frequent ones.

Keywords Datamining · Data streams · Itemsets · Minimum support · Sliding window

1 Introduction

The aim of this study is to generate frequent patterns from transactional data streams. The elements of a transactional data stream are itemsets. The frequent patterns are in terms of frequent itemsets. For example, a data stream of tweets posted by users where each tweet is an element of the data stream and is an itemset of words. The frequent patterns are the sets of words that occur in most of the tweets. The significance of a frequent pattern, in this case, is that it contains words related to a very significant event that day.

The recent advancement in hardware and software has resulted in the generation of enormous amounts of data leading to the concept of data streams [5]. A data

S. B. Naik (✉)
Sant Sohirobanath Ambiye Government College, Pernem, Goa, India
e-mail: xekhar@rediffmail.com

J. D. Pawar
DCST, Goa University, Taleigão, India
e-mail: jdp@gmail.com

A. J. Kulkarni et al. (eds.), *Proceedings of the 2nd International Conference on Data Engineering and Communication Technology*, Advances in Intelligent Systems and Computing 828, https://doi.org/10.1007/978-981-13-1610-4_36