Journal of Network and Computer Applications 79 (2017) 1–24

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

Review

Towards next-generation heterogeneous mobile data stream mining applications: Opportunities, challenges, and future research directions

Muhammad Habib ur Rehman[a,*], Chee Sun Liew[a], Teh Ying Wah[a,*], Muhammad Khurram Khan[b]

[a] Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia
[b] Center of Excellence in Information Assurance, King Saud University, Riyadh, Saudi Arabia

ARTICLE INFO

Keywords:
Frequent pattern mining
Classification
Clustering
Mobile computing
Cloud computing
Edge computing

ABSTRACT

The convergence of Internet of Things (IoTs), mobile computing, cloud computing, edge computing and big data has brought a paradigm shift in computing technologies. New computing systems, application models, and application areas are emerging to handle the massive growth of streaming data in mobile environments such as smartphones, IoTs, body sensor networks, and wearable devices, to name a few. However, the challenge arises about how and where to process the data streams in order to perform analytic operations and uncover useful knowledge patterns. The mobile data stream mining (MDSM) applications involve a number of operations for, 1) data acquisition from heterogeneous data sources, 2) data preprocessing, 3) data fusion, 4) data mining, and 5) knowledge management. This article presents a thorough review of execution platforms for MDSM applications. In addition, a detailed taxonomic discussion of heterogeneous MDSM applications is presented. Moreover, the article presents detailed literature review of methods that are used to handle heterogeneity at application and platform levels. Finally, the gap analysis is articulated and future research directions are presented to develop next-generation MDSM applications.

1. Introduction

The escalation in mobile data was witnessed about 4000-fold over the past decade (Cisco, 2015). Cisco, the big name in network infrastructures, predicts that mobile data will grow up to 30.6 Exabytes (*i.e.* 30.6 billion Gigabytes) by the year 2020 (Cisco, 2015). This massive amount of data will be generated by next generation of mobile systems such as mobile IoTs, WSNs, BSNs, robotics, unmanned aerial vehicles, and satellite systems to name a few (Rehman et al., 2016a). Considering this growth, mobile data will challenge the storage and processing capacities of existing computing systems. Next-generation applications will be developed to handle the data in streaming mode and on-the-fly using in-memory data processing architectures before storing in large scale distributed systems (Zhang et al., 2015). These trends will highlight the importance of data stream mining applications which perform in-memory analytic operations over streaming data in order to uncover hidden knowledge patterns (Krishnaswamy et al., 2012). These knowledge patterns will help understanding the underlying data and benefit in decision making in personal and commercial applications.

Mobile streaming data which is the subset of overall big data is helpful in improving business operations across the enterprises (Rehman et al., 2016b). For example, the analysis of mobile data streams generated by remote vehicles help in optimizing supply chain management operations (Kargupta, 2016). Similarly, the mobile data streams collected from remote customers is useful for creating personalized services for online shoppers (Tan et al., 2016). The governments can also improve the daily and emergency response management operations by analyzing real-time mobile streaming data from citizen's mobile devices (Murphy, 2016). Despite wide applicability, it is quite challenging to decide about where and when to process the streaming mobile data.

This article presents a thorough literature review of existing MDSM applications and platforms in order to establish the state of the art and find the future research directions. A few relevant literature reviews were proposed in the past, however, they emphasized on other perspectives. For example, the authors in Gaber et al. (2005), Parthasarathy et al. (2007), Goel et al. (2010), Fuqiang (2011), Krishnaswamy et al. (2012), Tsai et al. (2014), Gaber et al. (2014a), Nguyen et al. (2015) and Chen et al. (2015) focused on general MDSM algorithms and lack the discussion on application-level and platform-level issues. Similarly, in our previous study (Rehman et al., 2015), we

* Corresponding authors.
E-mail addresses: mhrehman@siswa.um.edu.my (M.H.u. Rehman), csliew@um.edu.my (C.S. Liew), tehyw@um.edu.my (T.Y. Wah).

http://dx.doi.org/10.1016/j.jnca.2016.11.031
Received 26 July 2016; Received in revised form 13 November 2016; Accepted 28 November 2016
Available online 01 December 2016
1084-8045/

studied mobile data mining applications in batch mode execution and static datasets. To the best of our knowledge, this is the first article that presents the review of MDSM applications and platforms in MECC environments. The article is structured as follows. Section 2 presents the bibliometric analysis of mobile data mining and mobile data stream mining publications which were indexed in web of science databases. Section 3 presents a detailed discussion on execution platforms for MDSM applications and the associated opportunities and challenges. Section 4 presents the taxonomy of heterogeneous MDSM applications. Section 5 presents a thorough literature review of methods for handling heterogeneity in MDSM applications. Section 6 discusses the heterogeneity issues at platform level and Section 7 presents a detailed literature review of selected platforms for MDSM applications. Section 8 presents the gap analysis of existing literature and discusses the future research directions. Finally, the article concludes in Section 9.

2. Bibliometric analysis of Web of Science databases

Research on mobile data mining is growing rapidly in recent years. We performed a preliminary study on Web of Science (WoS) databases (Web of science databases, 2016) by querying the string "mobile data mining". According to retrieved statistics, as of 28th January 2016, the WoS databases indexed 1930 publications in last 27 years (from 1990 to 28th January 2016) from International Scientific Indexing (ISI)-listed journals, conferences and workshop proceedings, and magazines (See Fig. 1). There was no significant research on the topic from 1990 to 2002. Since Year 2002, the miniaturization of technologies and on-board sensing technologies had geared-up the research on mobile data mining. However, the major boom started from Year 2007 when both Google (Android (operating system), 2016) and Apple (Apple iphone history, 2016) released their mobile operating systems.

According to Fig. 1, the number of publications rapidly increased till 2015 which shows that mobile data mining is continuously becoming a hot research topic. In near future, we perceive a major shift towards the research on mobile data mining due to rapid growth in far-edge mobile devices for example smartphones, wearable devices, mobile IoTs, and body sensor networks to name a few. The citation trends for the topic "mobile data mining" are depicted in Fig. 2. The citation analysis showed that publications on the topic of mobile data mining obtained 9041 total citations from 8180 citing publications which were indexed in WoS databases. The popularity of research on mobile data mining is witnessed by the fact that 7935 citing publications were published without self-citations by the respective authors. The average citations per publication is 4.68 with h-index as 40. Fig. 2 also depicts that arrival of mobile operating systems in 2007, boomed the research on mobile data mining and it is still increasing day by day.

Since the main focus of this article is on mobile streaming data, therefore, we further analyzed the bibliographic records from WoS databases with another query string as "mobile data stream mining". We found 112 publications indexed by WoS databases from Year 1990 to 28th January 2016. These 112 publications were cited by 343 other publications in WoS databases whereby 331 publications do not contain any self-citation by respective authors. The average citation per publication is 3.06 with h-index as 11 which was lower when compared with bibliometric analysis of "mobile data mining" because less number of publications on the topic. Likewise, the major boom in "mobile data stream mining" was also witnessed after Year 2007 and it is rapidly growing. Considering the fast growth of research in MDSM algorithms, applications, and execution platforms, a thorough literature review is presented in this article.

3. Mobile data stream mining platforms

The MDSM platforms facilitate in efficient execution of analytic components. The literature review reveals that MDSM platforms (see Fig. 3) were deployed in multiple topological settings (Abdallah et al., 2015; Gaber et al., 2014b; Haghighi et al., 2013; Jayaraman et al., 2014a). The underlying communication models include multiple computing devices and systems having different form factors. These devices and systems include mobile devices, Internet, and intranet based application servers and cloud data centers to name a few (Jayaraman et al., 2014b; Kargupta et al., 2010; Mukherji et al., 2014). The topological settings of MDSM platforms that are presented in this article are based on far-edge mobile devices, far-edge to far-edge communication models, mobile and immobile edge servers based communication models, mobile cloud computing and mobile edge cloud computing systems.

3.1. Far-edge mobile devices

Far-edge mobile devices are defined as any portable system or device with wireless communication interfaces and ability to produce or process data. Smartphones, wearable sensors, wireless body sensor networks, smart vehicles, and Mobile Internet of Things (IoTs) are a few examples of far-edge mobile devices. Although modern far-edge mobile devices enable rich MDSM applications such as virtual reality, computer vision, and multimedia applications using cloud augmented computational resources (Satyanarayanan et al., 2015) however the execution of heterogeneous MDSM applications inside far-edge devices is a challenging task (Rehman et al., 2015). Far-edge mobile devices

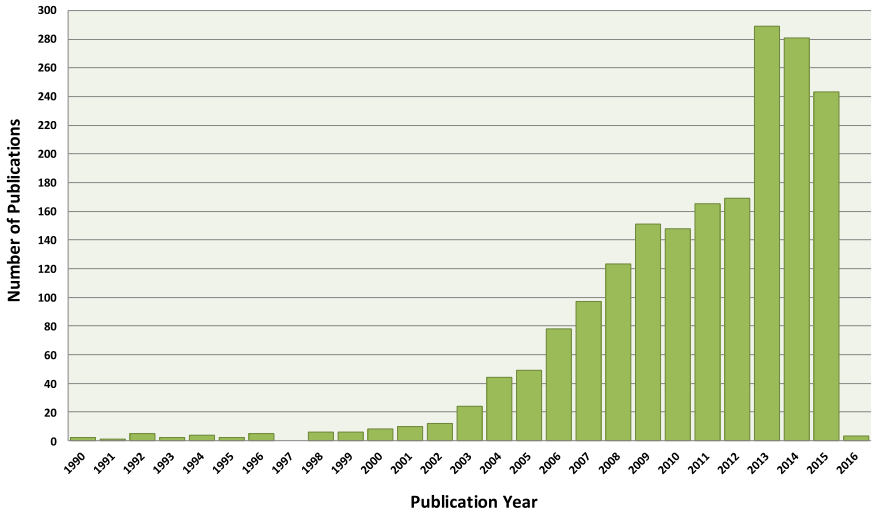

Fig. 1. Year-wise publications (1990–2016).