

G-SPAMINE: An Approach to Discover Temporal Association Patterns and Trends in Internet of Things

Shadi Aljawarneh^a, Vangipuram Radhakrishna^b, Puligadda Veereswara Kumar^{c,d}, Vinjamuri Janaki^c

^a Software Engineering Department, Jordan University of Science and Technology, Irbid, Jordan

^b Department of Information Technology, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India

^c Former Professor, Computer science and Engineering, University College of Engineering, Osmania University, Hyderabad, India

^d Department of Computer Science and Engineering, Acharya Institute of Technology, Bangalore, India

^e Department of Computer Science and Engineering, Vaagdevi Engineering College, Warangal, India

Abstract—Temporal data is one of the most common form of data in internet of things. Data from various sources such as sensors, smart phones, smart homes and smart vehicles in near future shall be of temporal nature with generated information recorded at different timestamps. We call all such data as time stamped temporal data. Discovery of temporal patterns and temporal trends from such temporal data requires new algorithms and methodologies as most of the existing algorithms do not reveal emerging, seasonal and diminishing patterns. In this paper, the objective is to find temporal patterns whose true prevalence values vary similar to a reference support time sequence satisfying subset constraints through estimating temporal pattern support bounds and using a novel fuzzy dissimilarity measure. We name our approach as G-SPAMINE. Experiment results show that G-SPAMINE out performs naive and sequential approaches and comparatively better to or atleast same as SPAMINE. In addition, the stamped temporal data adds extra level of privacy for temporal patterns in the IoT.

Keywords—temporal data; seasonal pattern; support bounds; temporal trend; web of things

I. INTRODUCTION

The term Internet of things is gaining significant importance in the present day and is surely going to dominate in future, the so called smart world which will soon be called “web of things”. The data generated in IoT can be from NFC or RFID Tags, devices, Sensors, Smart homes and vehicles, and any other Smart environments. IoT is a network of Smart of things in a simple sense. Currently, there is no unique global eco system of things available readily for IoT and hence it is virtually impossible even today. This is because there is no single protocol that is universally accepted and unique in context of IoT. Internet of Things as of today must be viewed as an Isolated Intranet of things and not as a single global network of Internet of things. As there is no single globally applicable protocol which can be practiced currently; the idea is to reuse the existing protocols and those others available which are used for coming up with flexible, interactive and scalable applications as practiced in the context of Web. It is here, where we come across the term, “**Web of Things**”, in perspective of IoT. In short, the idea of Web of things is to essentially, “Reuse and take advantage of already available, proved standards and protocols of web, so as to make, all the

data services and other application services become seamlessly accessible to a preponderant pool of web developers. Figure.1 shows the comparison of internet of things and web of things, [26].

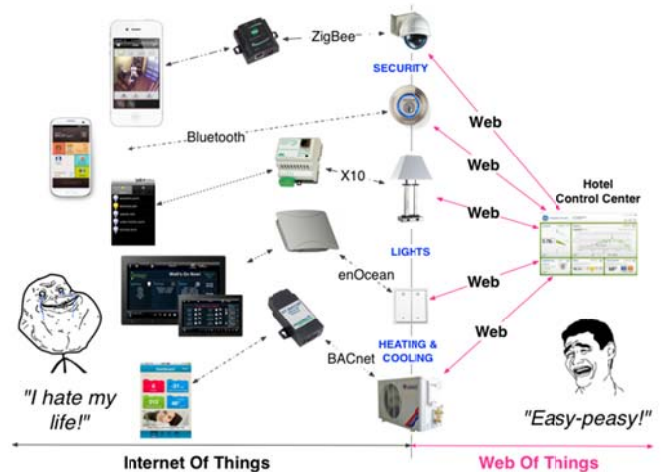


Figure 1. Internet of Things Vs Web of Things

Temporal data is one of the most common forms of data in the internet of things. Data from various data sources such as sensors, smart phones, smart homes and vehicles in near future shall be of temporal nature with generated information recorded at different timestamps. Such data may be called as time stamped temporal data. Discovery of temporal patterns, studying and knowing important trends in such time stamped temporal data requires new algorithms and methodologies as the existing algorithms do not reveal emerging, seasonal and diminishing patterns. Also the efficiency of computing resources is of prime concern. A traditional database system such as RDBMS is not suitable of handling time series data.

For example, consider the example of a power distribution company in INDIA which has millions of customers and users. The utility of each of these users is recorded day by day and stored as one reading per month. If the total number of users is 10 million, then the number of readings becomes 10 million per month. If we view it for a year, this comprises of 120 million readings per year, when we store it in a standard