IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 08, Issue 11 (November. 2018), ||V (I) || PP 11-14

A Methodology of Human Behavioral Pattern Mining Approach to Big Data Analytics

Dr.P.Srimanchari[1], Dr.G.Anandharaj[2]
[1]Assistant Professor and Head, Department of Computer Applications, Erode Arts and Science College(Autonomous), Erode – 638001
[2]Associate Professor and Head, Department of Computer Science, Adhiparasakthi College of Arts and Science (Autonomous), Kalavai, Vellore - 632506
Corresponding Author: Dr.P.Srimanchari

Abstract: This work introduces a set of scalable algorithms to identify patterns of human daily behaviors. These patterns are extracted from multivariate temporal data that have been collected from smartphones. We have exploited sensors that are available on these devices, and have identified frequent behavioral patterns with a temporal granularity, which has been inspired by the way individuals segment time into events. We briefly introduce the basics of related research topics, review state-of-the-art approaches, and present some preliminary thoughts on future research directions. This paper proposes a fusion of three different data models like relational, semantically, and big data based data and metadata involving their issues and enhanced capabilities.

Keywords: Multivariate temporal data, Big-data, real-time analytics.

Date of Submission: 08-11-2018 Date of acceptance: 22-11-2018

I. INTRODUCTION

The computing and networking capabilities of mobile and wearable devices, makes them appropriate tools for obtaining and collecting information about user activities (mobile sensing). This has led to a significant expansion of opportunities to study human behavior ranging from public transport navigation [1] to well-being [2]. Moreover, the advent of mobile and wearable devices enables researchers to unobtrusively identify human behavior to an extent that was not previously possible. Nevertheless, there is still a lack of wide acceptance of mobile sensing applications in real-world settings [3]. There are different reasons for this mismatch between capability and acceptance. First, the limitation of resources and a lack of accuracy in the collected contextual data, especially is a challenge with regard to the battery life [4]. Furthermore, the small size of sensors that are dealing with radio frequency, i.e., Bluetooth, WiFi and GPS, affects the quality of their data [5] (the smaller the device, the less accurate the data).

Big data in real time have diverse and autonomous representations bringing highly unstructured and unrelated data based relationships in producing results which are getting complex and faulty. The heterogeneous data features represent different representations for data. Decrease the effect of heterogeneous and complex data; there can be computationally introduced at localized systems considering they are having better computational power. There can be a way of transforming data into a common data fusion. As a result, the common forms of data in consequence to data fusion will be highly compatible for data linkage and relativity indexing for getting better analytical outcomes. Major of data is stored either using relational, semantical or big data formats. Relational data is stored in the form records containing a collection of singleton cells representing fields supported by its data structure and constraints for an entity.

II. RELATED WORKS

The relational data model was first invented with the term``relational database" by E. F. Codd from IBM in 1970.Whereas, Codd had defined relational in his paper titled``A Relational Model of Data for Large Shared Data Banks"in which he had introduced 12 rules for implementing relationaldata model also known as Codd's rules. These ruleswere completely taken but up to a minimum and necessarylevel in defining a table as a relation and operatorsused to manipulate this data form. Whereas, a languagewas introduced for querying by Chamberlin and Boycein 1974 from IBM. It was first named a SEQUEL (StructuredEnglish Query Language) which was made standard in ANSIX3H2 committee with SQL (Structured Query Language)in 1986 [9]. In 1976 a designing model to view relationaldata with the entity-relational model by Peter Chan. In 1990'sthird generation database system manifesto was introducedby Stonebraker in 1990 which in 1996 became ORDBMS

(Object Relational Databases Management System) [10].Further history of RDB is concerned with the managementsystem of the relational model. The evolution of RDB data and querying model linkingthem together according to the timeline at the side to showtheir arrival according to the history using year and authordetails. Now in next section evaluation of XML is beingrepresented [13], [14].

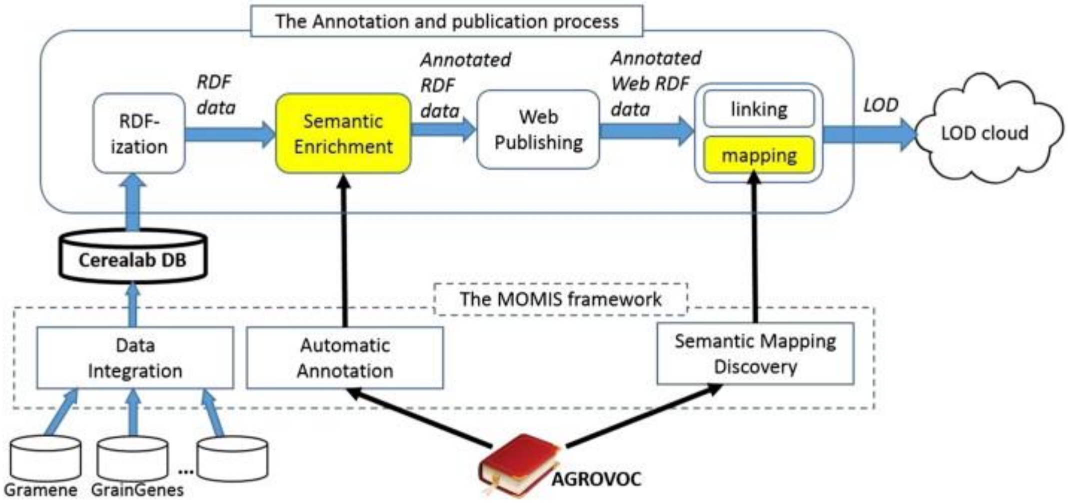


Figure 2.1 The annotation and publication process

Inventory Forecasting Use Case: Inventoryforecasting is one of the major tasks in manufacturingand includes raw material inventory forecasting.Accurate and reliable inventory prediction canguarantee a smooth production process[16]. Herewe assume that the forecasting target is the inventoryof a PCB (Printed Circuit Board) and the address of thetraining data is inventoryOfPCB.csv.Car Evaluation Use Case: When manufacturingenterprises are going to launch a new product, itneeds to match the public's aesthetic. We can usea rule extraction model to extract rules regarding thepublic's aesthetic that were hidden in previous userfeedback data on various car models. Here we takea Car Evaluation Data Set from the UCI machinelearning repository as training data[17]; the address iscarEvaluation.csv and the target is carAcceptability.Tool Condition Monitoring Use Case:Manufacturing systems are becoming morecomplex and are subject to failures that adverselyimpact their reliability, availability, safety, andmaintainability[18]. For example, in the high-speedmilling process, a worn milling tool might irreversiblydamage a workpiece[19]. In such a case, real-time monitoring of the condition of the tools can help theoperator avoid catastrophic events. Here we take a SteelPlates Faults Data Set as training data[20]; the address issteelPlatesFaults.csv and the target attribute is Faults.

III. EVENT MODEL DESCRIPTION

Asthe interface for the whole system, GMDL is one ofits most essential parts. It directly determines whattasks GMDA can perform and thus is crucial to thewhole system, therefore it requires perfect design. Weattempt to make it concise and easy to understand forinexperienced users.

3.1 GMDL for use cases

To fully describe a task, we must specify at least threekey points:
The goal of task;
The dataset;
Target attribute (if the task is prediction).
Each point is independent, and to make GMDAeasier for inexperienced users, we initially propose abasic formal structure for the GMDL language, whereCOMMAND is the description of the task that GMDAcan understand. That is:

COMMAND -> SETTING COMMAND | SETTING;
SETTING -> PROPERTY = "VALUE ".

3.2 Definition of language GMDL

COMMAND includes multiple independent SETTINGs, which are independent because theyare assignments for different PROPERTYs.

3.3 Annotation for Big Data

Real-time data collection is found mostly in the form of sensorsdata collected through physical or biological resources.In the current era of information analytics Internet ofThings (IoT) is playing the main role in managing, controllingand monitoring of the resources even at remote locations.With the involvement of social medium and mobilecommunication data is increasing rapidly. At the end of bigdata, Hadoop is playing a key role through its platform indata collection, computational clustering of distributed units,and dramatic fast analytics. However, still, it lacks in realtimeboosted analytics for a localized fast outcome. Forthat to work data fusion is proposed at the level of localizedor short area cluster of units to have highly interactive.

3.4 Understanding Challenges

The challenges involved in the methodology for real-timedata fusion for localized big data's analytics concerns withdata updates. Other issues involve one data model supportand limitation to other data model during the process of datafusion. Data collected in traditional data storage representingrelation database where data is placed separately frommetadata. The new generation data formats like, JSON andRDF are more data and hierarchy oriented.

IV. COMMUNICATION PROTOCOLS

In order to implement our algorithms for the problemdescribed above, first the data format should be converted from heterogeneous data to machine-processable data, i.e.,the raw data needs to be converted to the previously described entity format. As previously stated, the data hasbeen collected from heterogeneous sources. Some sensorshave multiple values, for instance WiFi has BSSID, SSIDand Capalities (WPA, PSK, etc.). Nevertheless, for each sensorour model chooses only one value. In particular, eachsensor (attribute) A, requires a single data point (value) D.Therefore, “BSSID” has been chosen for WiFi and Bluetooth,the pseudonymized phone number for SMS and Calls,“process name” for Application and tilting, in-vehicle, on-bicycle, walking, still, and unknown for the activity sensors(UbiqLog uses Google play services for activity recognitionand therefore there is no raw accelerometer datainside the dataset). A similar approach has also been usedfor the Device Analyzer dataset, which we do not report ithere to preserve space.During the second step, we propose an algorithm thatidentifies the movement (based on location changes) state,which will be used to enrich the semantics of the data withinthe notion of the location. In third step, we need to convertthe timestamp to a time similar to the human perception oftime. Afterward, in the fourth step we describe the behaviorsimilarity and FBP detection algorithms.

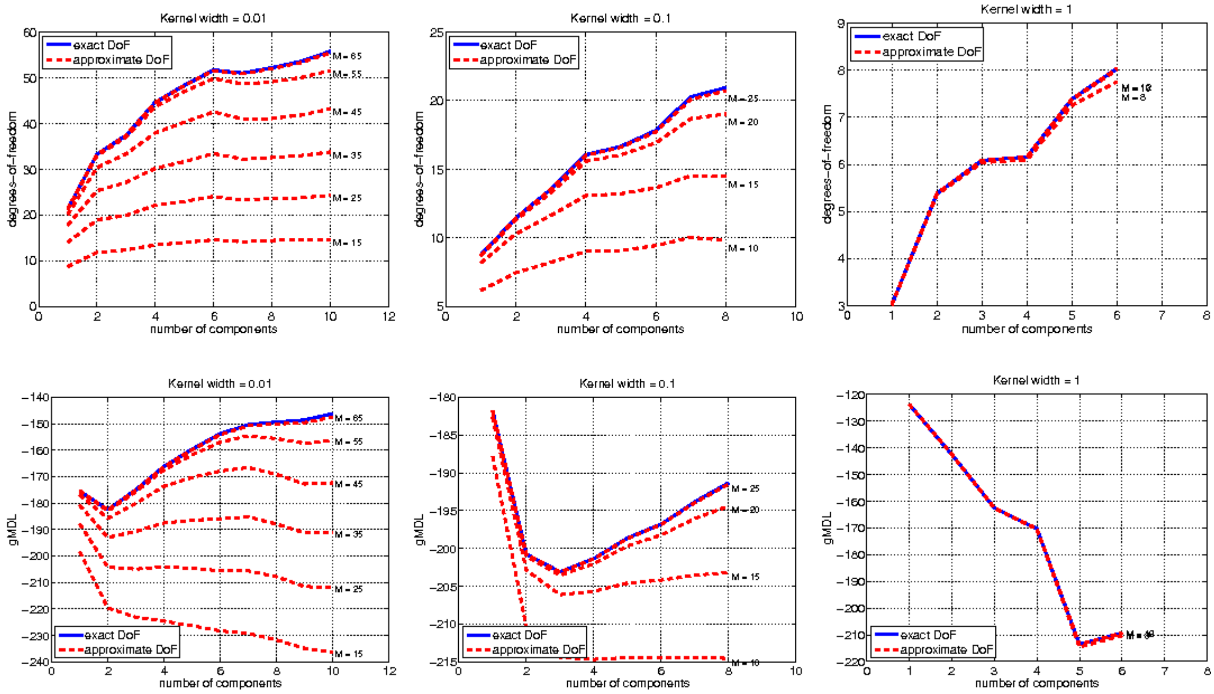
V. SIMULATION

5.1 Accuracy Analysis -Ground Truth Dataset

In order to evaluate the accuracy and quality of the identifiedFBPs, we have created a ground truth dataset, which is composedof more than 5,000 identified entities (that participatein FBPs), from five users. It contains randomly identifiedFBP data that has been labeled by the users as either true orfalse.

5.1.2 Accuracy of Identified FBPs

After collecting the labels, we carefully examined the accuracyof our algorithms using three temporal segments of theday: 0:00-07:59 (0-8), 08:00-15:59 (8-16) and 16:00-23:59 (16-24) and different TGs. This time based segmentation hasbeen inspired by similar work in mobile data mining,but it is more accurate than the two divisions proposed byMa et al.



VI. CONCLUSION

We have proposed a scalable approach fordaily behavioral pattern mining from multiple sensor information.This work has been benefited from two real-worlddatasets and users who use different smartphone brands.We use a novel temporal granularity transformation algorithmthat makes changes on timestamps to mirror thehuman perception of time. Our frequent behavioral patterndetection approach is generic and not dependent on a singlesource of information; therefore, we have reduced the riskof uncertainty by relying on a combination of informationsources to identify frequent behavioral patterns.

REFERENCE:

[1] P. Y. Vandenbussche, G. A. Atemezing, M. Poveda-Villalón, and B. Vatant, ``Linked Open Vocabularies (LOV): A gateway to reusable semantic vocabularies on theWeb,'' *Semantic Web*, vol. 8, no. 3, pp. 437_452, 2017.

[2] K. Höffner, ``Survey on challenges of question answering in the semantic Web,'' *Semantic Web*, vol. 8, no. 6, pp. 895_920, 2017.

[3] M. Nentwig, M. Hartung, A.-C. N. Ngomo, and E. Rahm, ``A survey of current link discovery frameworks,'' *Semantic Web*, vol. 8, no. 3, pp. 419_436, 2017.

[4] Gangemi, V. Presutti, D. R. Recupero, A. G. Nuzzolese, F. Draicchio, and M. Mongiovì, ``SemanticWeb machine reading with FRED,'' *Seman- tic Web*, vol. 8, no. 6, pp. 873_893, 2017.

[5] R. Goodman, R. P. Mahler, and H. T. Nguyen, *Mathematics of Data Fusion*, vol. 37. Dordrecht, The Netherlands: Springer, 2013.

[6] M. Bevilacqua, A. Tsourdos, A. Starr, and I. Durazo-Cardenas, ``Data fusion strategy for precise vehicle location for intelligent self-aware maintenance systems,'' in *Proc. IEEE 6th Int. Conf. Intell. Syst., Modelling, Simulation (ISMS)*, Feb. 2015, pp. 76_81.

[7] Hotho, R. Jäschke, and K. Lerman, ``Mining social semantics on the social Web,'' *Semantic Web*, vol. 8, no. 5, pp. 623_624, 2017.

[8] D. Calvanese*et al.*, ``Ontop: Answering SPARQL queries over relational databases,'' *Semantic Web*, vol. 8, no. 3, pp. 471_487, 2017.

[9] R. Mukherjee, ``Interfacing data destinations and visualizations: A history of database literacy,'' *New Media Soc.*, vol. 16, no. 1, pp. 110_128, 2014.

[10] G. Palmer, P. A. Stephens, A. I. Ward, and S. G. Willis, ``Nationwide trophic cascades: Changes in avian community structure driven by ungulates,'' *Sci. Rep.*, vol. 5, Oct. 2015, Art. no. 15601.

Dr.P.Srimanchari. " A Methodology of Human Behavioral Pattern Mining Approach to Big Data Analytics.” IOSR Journal of Engineering (IOSRJEN), vol. 08, no. 11, 2018, pp. 11-14.