Future Generation Computer Systems ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Future Generation Computer Systems

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

An integrated information lifecycle management framework for exploiting social network data to identify dynamic large crowd concentration events in smart cities applications

George Kousiouris [a,*], Adnan Akbar [b], Juan Sancho [c], Paula Ta-shma [d], Alexandros Psychas [a], Dimosthenis Kyriazis [a], Theodora Varvarigou [a]

[a] Department of Electrical and Computer Engineering, National Technical University of Athens, 9, Heroon Polytechniou Str, 15773 Athens, Greece
[b] G Innovation centre (5GIC), Institute for Communication Systems, University of Surrey, Stag Hill Campus, Guildford, UK
[c] ATOS Research & Innovation, Internet of Everything Lab, Av. Capuchinos Basurto, 6, Bilbao, Spain
[d] IBM Research, University of Haifa Campus, Mount Carmel, Haifa, 3498825, Israel

HIGHLIGHTS

- Integrated and automated framework including Smart City platform, Twitter and Analytics services.
- Middleware layer for adaptation and synchronization needed.
- 2 month experiment in the city of Madrid.
- Fine grained crowd concentration identification per location is feasible via Twitter data.
- Optimal thresholds can ignore timeslots in this case and be based on overall CDF limits.

ARTICLE INFO

Article history:
Received 7 January 2017
Received in revised form 31 May 2017
Accepted 10 July 2017
Available online xxxx

Keywords:
Internet of things
Social networks
Analytics
Cloud computing
Event identification
Smart cities

ABSTRACT

With the current availability of an extreme diversity of data sources and services, emerging from the Internet of Things and Cloud domains, the challenge is shifted towards identifying intelligent, abstracted and adaptive ways of correlating and combining the various levels of information. The purpose of this work is to demonstrate such a combination, on one hand at the service level, through integrating smart cities platforms for user level data, and on the other hand at Complex Event Processing, Storage and Analytics capabilities together with Twitter data. The final goal is to identify events of interest to the user such as Large Crowd Concentration (LCC) in a given area, in order to enrich application level information with related event identification that can enable more sophisticated actions on behalf of that user. The identification is based on observation of Twitter activity peaks compared to historical data on a dynamic time and location of interest. The approach is validated through a two-month experiment in the city of Madrid, identifying LCCs in sporting events around two sports venues and analyzing various approaches with relation to the needed thresholds definition.

© 2017 Published by Elsevier B.V.

1. Introduction

In the current technological landscape, the major challenge is to combine versatile data sources in an intelligent manner, integrating and reasoning in order to infer situational awareness. Thus one can transcend from the level of raw data to the level of knowledge and wisdom (according to the DIKW pyramid [1]). Especially with the advent of the Internet of Things, these data sources are expected to reach on an unprecedented scale [2], therefore enabling the optimization of multiple domains of an individual's day to day activities [3]. Among these domains, applications in transport such as alerts to passengers, smart cities transport, crowd and traffic management are among the key identified aspects.

However, one of the major challenges is the ability to integrate these sources and to reason in order to exactly harvest the added value from this large availability of data [2,4]. On one hand a clear and repeatable process needs to be defined in order to link the different elements of such a system, catering for a decoupled integration approach, while a specific added value and context needs to be derived by each case specific analysis of the data that might prove useful to the recipients at the application layer. The need for

* Corresponding author.
E-mail address: gkousiou@mail.ntua.gr (G. Kousiouris).

http://dx.doi.org/10.1016/j.future.2017.07.026
0167-739X/© 2017 Published by Elsevier B.V.

Please cite this article in press as: G. Kousiouris, et al., An integrated information lifecycle management framework for exploiting social network data to identify dynamic large crowd concentration events in smart cities applications, Future Generation Computer Systems (2017), http://dx.doi.org/10.1016/j.future.2017.07.026.

a concise and multi-source big data and analytics framework for addressing current challenges is critical [5].
In order to design and implement such an approach, a set of further fine grained requirements need to be defined:

- Flexibility in terms of incorporating diverse interfaces and protocols in order to communicate between the different subsystems of different nature and scope, thus providing the combinatorial nature of data analysis and correlation from multiple sources (**Req. 1**)
- The produced information should be asynchronously sent in a notification manner, so that there is dynamicity in the receipt of information only for the case of interest (**Req. 2**)
- A fully automated and dynamic process should exist in order to be able to define arbitrary points of interest (**Req. 3**)
- The analysis performed based on the previous point should take under consideration the context of the given point (e.g. in terms of the specific location or timeslot of interest), adapting to each case in a fine-grained manner (**Req. 4**)
- The implementation should be offered as a service, API or other pluggable form for maximum flexibility and decoupling of implementation details (**Req. 5**)

The aim of this work is to present such an approach that includes the following main points:

- Integration between smart city data, coming from a specific passenger monitoring system in the city of Madrid (Reactivebox [6]) and social network data coming from Twitter, for alerting passengers with special needs (e.g. children, elderly, persons with cognitive restrictions etc.) and their caregivers about Large Crowd Concentrations (LCC) along their journey. This will aid them in avoiding confusion and reduced mobility circumstances when getting off at an area where an unexpected amount of people is concentrated. Novelty of the contribution in this case is inserted through the usage of a general purpose data source (such as Twitter data) for reasoning and inference on the state of a city region through an approach that does not require extensive knowledge of e.g. natural language processing or other complex approach. 5 different approaches of threshold investigation are identified and analyzed in terms of their accuracy and fit-for-purpose, based on the number and type of errors. Another aspect of novelty is the combination and fusion of information from one application domain to another in a cross-verticals integration, that demonstrates the value of diffusing knowledge externally to its main domain of usage, thus minimizing silos of data and exploiting multi-source data correlation. Thus a higher level of cognition and knowledge may be achieved, indicating the way towards numerous others combinations (and more importantly, state of mind to perform the combinations) that may be performed in the context of Smart City applications.
- Ingestion of the integrated data in scalable Cloud based solutions (Openstack Swift) integrated with analytics tools (Apache Spark) for directly working on the acquired datasets and Complex Event Processing tools in order to monitor and issue alerts from streaming data. The solution is able to adapt to heterogeneous data by adjusting the specified Apache AVRO template used for annotation. Novelty of the contribution in this case is represented through the use of specialized tools per case, which enables the exploitation of each tool's powerful features and focus on a specific domain, instead of general purpose tools that could ease integration aspects but would come in the cost of reduced functionality. Furthermore a contribution of this work is the coordination of a large part of the process (in the knowledge extraction and on-line identification) through a graphical, web based tool like Node-RED which may enable multiple roles to interact with the produced system, since it minimizes the entry level knowledge needed for each layer and gives the ability to interact based on message formats and production/consumption of event information. Furthermore, the created flows can be easily copied and adapted to new cases of interaction.
- Connection of all these systems with an application based logic, achieved through a middleware layer based on Node-RED, in order to orchestrate the necessary actions in the foreseen data flow and provide the necessary adaptations in terms of protocols and data formats. Novelty of the contribution in this case is represented through the successful usage of Node-RED, a tool primarily used for interconnection in the IoT domain, as a general purpose integration and application logic mechanism, exploiting its powerful abstraction and intuitive usage features to speed up development and integration, while adapting to a multitude of protocols (indicatively DDP, MQTT, AMQP, REST, SQL), asynchronous, event-driven logic, push and pull acquisition models and different data formats (XML, JSON, and overall 4 different data schemas). Thus a contribution of this work is the proof of concept that this tool can be used also as a middleware layer in the context of Smart Cities, and not only at the level of Things.
- Fine grained analysis of a specific location needed by a specific user in order to identify a specific event such as a Large Crowd Concentration (LCC), through dynamic receipt and analysis of the respective data that can aid in this identification. The novelty of the contribution in this case is that there is no static division of the city in regions but in each case the adaptation is performed dynamically with relation to the specific user's route and the specific limited area of interest (such as the drop-off point of the journey). Such fine grained analysis enables enhanced accuracy and optimal adaptation at the user level, individualized focus and personal scope, a key feature of Web 3.0 applications.

The relation between the DIKW pyramid, the various levels of information, the involved technology enabling layers and the mapping on Twitter messages processing are presented in Fig. 1. It is necessary to stress that while this LCC analysis is performed in the context of the specific application, its consumption may also be extended to other cases in which an LCC event could be the target of a specific action (e.g. police engagement and monitoring, marketing approach etc.), based each time on the scope under which the LCC event is consumed. Furthermore, the methodological aspects may be replicated in different types of events, by adjusting the baseline ingestion of data and extraction of knowledge.

The paper proceeds as follows. In Section 2 related work is presented with accordance to the usage of social networking data and the technologies used in this work, while Section 3 describes the application context and the general structure of the implemented prototype. Section 4 aims to validate the usage of the framework in an identification experiment including sports events as a source of indisputable gatherings and analysis of the related tweet patterns, while Section 5 concludes the paper.

2. Related work

Given that the matters investigated in this work are cross-domain, numerous technologies and approaches exist in each field. In this section the major points of interest will be investigated. What is common ground is the need for a repetitive and methodological approach with a clear focus on the middleware and

Please cite this article in press as: G. Kousiouris, et al., An integrated information lifecycle management framework for exploiting social network data to identify dynamic large crowd concentration events in smart cities applications, Future Generation Computer Systems (2017), http://dx.doi.org/10.1016/j.future.2017.07.026.

Download English Version:

https://daneshyari.com/en/article/4950189

Download Persian Version:

https://daneshyari.com/article/4950189

Daneshyari.com