A System for Traffic Events Detection using Fuzzy C-Means

Hayder Endo Pérez[1], Amed Leiva Mederos[1], Daniel Galvez Lio[1], Luis Ernesto Hurtado[1], Doymer García Duarte, and Ghislain Atemezing[2]

[1] Research Center in Informatics. Semantica Web Group. Central University "Marta Abreu" of Las Villas, Cuba
{hendo,amed,dgalvez,luhgonzalez,dgduarte}@uclv.cu
[2] Mondeca. Paris, France
ghislain.atemezing@mondeca.com

Abstract. Systems for traffic events administration are important tools in the prediction of disasters and management of that of the movement flow in diverse contexts. These systems are generally developed on non-fuzzy grouping algorithms and ontologies. However, the results of the implementation do not always give high precision scores due to different factors such as data heterogeneity, the high number of components used in their architecture and to the mixture of highly specialized and diverse domain ontologies. These factors do not ease the implementation of the systems able to predict with higher reliability traffic events. In this work, we design a system for traffic events detection that implements a new ontology called *trafficstore* and leverages the fuzzy c-means algorithm. The indexes evaluated on the fuzzy c-means algorithm demonstrates that the implemented system improves its efficiency in the grouping of traffic events.

Keywords: Semantic Web · Fuzzy C-means · IoT · Traffic Event Detection.

1 Introduction

The prediction of traffic events has been one of the most important research fields in smart cities. Sensors are an important source of data available in this field today. While sensor data can be published as mere values, searching, reusing, integrating, and interpreting this data requires a little more than the verification of observation results. Intelligent traffic systems use sensors to correctly interpret the values of humidity, carbon dioxide, etc. The sensors provide information on the observed properties and the sampling strategy used. The Open Geospatial Consortium (OGC) sensor web standards provide a means to annotate sensors and their observations. Nevertheless, these standards are not integrated or aligned with the W3C Semantic Web technologies and Linked Data in particular, which are key mechanisms for creating and maintaining a tightly interconnected global data graph. With the rise of the Web of Things (WoT), smart cities and smart buildings in general, actuators and the data produced by the sensors become first-class citizens of the web. One of the most used ontologies in the detection of traffic events is called SSN, which follows a horizontal and vertical modular architecture by including a lightweight but autonomous central ontology called SOSA (Sensor, Observation, Sample and Actuator) [17], which are key mechanisms for creating and maintaining a global and tightly interconnected graph of data. SSN/SOSA are used in several existing work in the literature to semantically annotate and analyze data in the IoT domain.

The development of traffic systems in smart cities goes through various development paradigms, among which is OBDA. These traffic management systems use OBDA (Ontology Based Data Access) that allows access to information stored in heterogeneous data sources through an abstraction layer that mediates between data sources and data consumers. Some of these tools used in OBDA are D2RQ, Mastro, morph-RDB, Ontop, OntoQF, Ultrawarp and others. In the OBDA paradigm, an ontology defines a high-level global schema of (existing) data sources and provides a vocabulary for user queries. An OBDA system rewrites such queries in both the ontologies and the vocabulary of the data sources and then delegates the actual evaluation of the query to a suitable system [8]. The DL-Lite description logic, which underpins OWL 2 QL is the one that represents the ontology in OBDA [15].

Traditional traffic management systems have evolved to Data Stream Management Systems [7, 11]. It is a system that is designed to execute continuous queries on a continuous flow of data or data stream. Traffic data is data that is generated from many sources, typically sending data records simultaneously in small-sized sets (several kilobytes).

This data must be processed sequentially and incrementally, record by record or in incremental time windows, and is used for a wide variety of analysis types, such as correlations, aggregations, filtering, and sampling. The information derived from the analysis provides the transport companies with the visibility of numerous aspects of the business and the activities of the clients, such as the use of the traffic service, the server activity, the clicks on the website of the transport company and the geographical location of means of transport, people and goods, and allows them to respond quickly to any situation that arises [6]. This would be great when it comes to detecting and correcting some events that traffic systems may have in advance.

IoT applications for traffic control in smart cities should have the ability to process event broadcasts in real time, extract relevant information and identify values that do not follow current trends. Beyond the identification of relevant events, the extraction of high-level knowledge from heterogeneous and multimodal data flows is an important component in traffic control systems. Existing flow reasoning techniques use prior knowledge and flow queries to reason about data flows. However, they do not meet the needs of IoT due to the lack of adequate treatment of uncertainty (for example, possible reasons for traffic jam) in the IoT environment [7]. Nor are they operable in any traffic environment due to the constant overlaps that generate the results of their services. This article seeks to provide with solutions to detect traffic events with fuzzy algorithms to achieve greater efficiency in transport event detection systems. The paper is organized as follows: Section 2 describes related work in the domain of streaming data and traffic events, followed by our proposal in Section 3 composed of an ontology, a system architecture. The first validation with results are shown in section 4. Section 5 provides some conclusions remarks.

2 Related work

Intelligent transport systems are a set of technological solutions used to improve the performance and safety of road, air and land transport. A crucial element for the success of these systems is the exchange of information, not only between vehicles, but also between other components of the road infrastructure through different applications. One of the most important sources of information in this type of system is the sensors. The sensors can be inside vehicles or as part of infrastructure, such as bridges, roads, traffic signs, traffic lights, etc. These sensors can provide information related to weather conditions and the traffic situation, which is useful to improve the driving process. To facilitate the exchange of information between different applications that use sensor data, a common knowledge framework is needed to enable interoperability.

In recent years, there has been a growing interest in ontologies for road transport systems. Gorender and Silva [9] in their work developed an ontology to represent road traffic. Their goal was to build a reliable traffic information system that provides information on roads, traffic, and vehicle-related scenarios on the highways. It also helps the traffic information system to analyze specific critical situations in this environment. For example, an ambulance may need to know the congestion status of a toll plaza. Requesting this information is essential if the ambulance is transferred to the scene of the accident. On the other hand, if a normal vehicle is moving down a highway without rushing, then this requested information is not that critical. Morignot and Nashashibi in [14] proposed a high-level representation of an automated vehicle, other vehicles, and their environment, which can help drivers make 'unorthodox' but practical relaxation decisions (for example, when a damaged car does not allow circulation, make the decision moving to another lane by crossing a solid line and passing the stopped car, if the other lane is clear). This high-level representation includes topological knowledge and rules of inference, in order to calculate the next high-level movement to be made by an automated vehicle, to aid the driver's decision-making. The main weakness of this approach is the lack of rules that represent the previous traffic regulations. They have just defined a set of traffic law infractions, that allow classifying the motion given as "legal" or "illegal". Zhao, IchiseySasakien [11] introduced an ontology-based knowledge base, containing maps and traffic rules. It can be aware of speed situations and make decisions at intersections to comply with traffic regulations, but it does not consider important elements such as traffic signs and weather conditions. The proposed work in [5] is an approach to creating a generic situation description for advanced driver assistance systems using logical reasoning in a traffic situation knowledge base. It contains multiple objects of different types such as vehicles and infrastructure elements such as roads, lanes, intersections, traffic signals, traffic lights and relationships between them. Logical inference is made to verify and expand the description of the situation and interpret the situation, for example, by reasoning about traffic rules. The capabilities of this ontological description approach are shown in the example of complex intersections with multiple roads, lanes, vehicles, and different combinations of traffic signals and traffic lights. As a restriction, in this work, the destination road that passes over the intersection must be known for each vehicle, so it is not possible to model different possibilities according to the actual situation of the intersection.

From an implementation point of view, there are several services whose practical applications already serve as services in some territories. Londonair is the London Air Quality Network (LAQN) website that shows air pollution in the City of London and South East England. The website provides information for the public, policy users, and scientists. LAQN was formed in 1993 to coordinate and improve air pollution monitoring in London. The network provides independent scientific assessments and measurements. This site does not monitor the entire area [13]. This site arises due to concerns regarding air pollution on the part of the London Department of Public Health. The measurements obtained here are used, in addition to evaluating air pollution, to track its trends over time and to create models that can evaluate how different government policies affect air pollution. These measurements also help to comply with the legal obligations of local authorities regarding air pollution. London air is maintained by the Environmental Research Group at Imperial College London. Monitoring is owned and funded by local authorities, Business Improvement Districts.

Open Data DK, as another solution, which is an association of Danish municipalities and regions that since 2016 have collaborated to open their data, that is, in a common open data portal [5]. Open data is non-personal data that anyone can access and use for free. They can be anything from data on municipal infrastructure to socioeconomic composition. The purpose of open data is: (i) Create transparency in public administration, (ii) Create fertile ground for data-driven growth and innovation and (iii) Ensure a higher degree of utilization of the data already collected. In this way, open data can be used in the development of applications and services or be a starting point for analysis, trend assessments, research, etc. Open Data DK has a data portal from which everyone can use the data for free without registering. The data portal is based on the open source software CKAN (Open Knowledge Foundation). All the datasets in the Open Data DK portal are grouped together to create an overview. They are divided into the same categories as in the European Data Portal. Among some of these datasets that are offered are those related to:

- Population and society, which include: demographic data, migration, employment, socioeconomics, etc.
- Energy, which include: energy consumption, energy sources, etc.

– Health, which include: assisted devices, dental care, drug / alcohol treatment, etc.
– Transport, including: mobility, parking, roads, winter maintenance, public transport, etc. such as agriculture, fishing and rural communities in the UK.

3 System Design

In this section, we describe the ontology for stream annotation, the system architecture, our annotation system and the different formats flowing from different modules of the system.

3.1 Ontology

The development of this system involves the management of a new ontology. As mentioned above regarding the reuse of ontologies, the information system adopts several concepts that are considered core attributes to provide a real-world context to the IoT stream and traffic events.

The ontology relates the spatial attributes of `IotStream`. The W3C geographical ontology provides a set of basic concepts that represent the location of a feature or a traffic event. The main concept of interest is the `geo:Point` that contains geospatial properties (latitude, longitude, and altitude). The IoT-lite ontology [2] expands properties to include relative location and relative altitude. To maintain the historical context of `StreamObservations`, especially in the case of mobility, a `geo:Point` can be linked to each `StreamObservation`. IoTStream is also associated with a defined coverage area where it is also relevant. Regarding the IoT-lite ontology: Coverage concept is used for simple coverage definitions, and GeoSPARQL [1], a well-established ontology for spatial attributes.

The next subset of concepts adopted relates to the IotStream generating source, the phenomena, and the measurement of your observations. As the streams in the real world are generated by sensors, the SOSA ontology has been used [10]. Through the object properties defined by *IoTLite*, the concepts `qu:QuantityKind` and `qu:Unit` of the QU ontology are also linked.

Although it is the sensor that generates the IoT stream, over the Internet, the stream data is usually provided by a TCP/IP application layer service. The IoT-Lite class provides `iot-lite:Service` class that contains fields related to the address of the service endpoint, the interface type, and the link to the interface description, which provides details on how to interact with the service.

Finally, over the lifespan of an IotStream, the quality of stream observations can change over time. For data analysis, knowledge of quality is very important so that adaptation measures can be applied when necessary. The Quality of Information (QoI) ontology provides the qoi:Quality concept that has subclasses that focus on a particular aspect of data quality, such as `qoi:Timeliness` and `qoi:Completeness` of observations. Table 1 lists the namespaces of the linked ontologies and their preferred prefixes.

Table 1. Prefixes and namespaces of the linked ontologies.

Prefix	Namespace
iot-lite	http://purl.oclc.org/NET/UNIS/fiware/iot-lite#
iot-stream	http://purl.org/iot/ontology/iot-stream#
owl	http://www.w3.org/2002/07/owl#
qoi	https://w3id.org/iot/qoi#
what	http://purl.oclc.org/NET/ssnx/qu/qu#
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs	http://www.w3.org/2000/01/rdf-schema#
sosa	http://www.w3.org/ns/sosa/
wgs84_pos	http://www.w3.org/2003/01/geo/wgs84_pos#
xml	http://www.w3.org/XML/1998/namespace
xsd	http://www.w3.org/2001/XMLSchema#
traff	http://sem.uclv.cu/def/trafficstore#

New classes were placed in the ontology with the prefix `traff` combining new conditions for `FeatureOfInterest` which describes the conditions that may affect any means of transport including `Climatologycall` Condition, `EnvironmentalCondition` and `TrafficConditions`. To monitor the traffic conditions, `SeaSection`, `RoadSection`, `LineSection` and `AirSection` classes were created. It is important to describe the Observation class where observations are detailed for plane, ship, train line and motor vehicle.
Likewise, the `TrafficConditions` class and the `EventCauses` class are detailed, in which causes of traffic problems are described, associated with international traffic rules and speed limits for various types of means of transport figure 1 and table 1[3].

3.2 System Architecture

There are many works related to this topic. This paper is dedicated to improving the architecture of the referenced system [6]. It has several components that can be optimized in the implementation process. The proposed architecture is simple and adapts to the conditions of the work environments. Following the work in [6] these components are reused: Producer, Analytics Service, Consumer and the Registry (See Figure 2).

1. Producer: The producer is responsible for fetching stream observations from the designated source of open data according to the application domain and publishing them to the message broker (4).
2. Analytics Service 1: The analytics service 1 is responsible for grouping atomic data points into windows of several sizes which may be consumed by another analytics service.

[3] The online documentation is available at `http://linkedvocabs.org/onto/trafficstore/trafficstore.html`

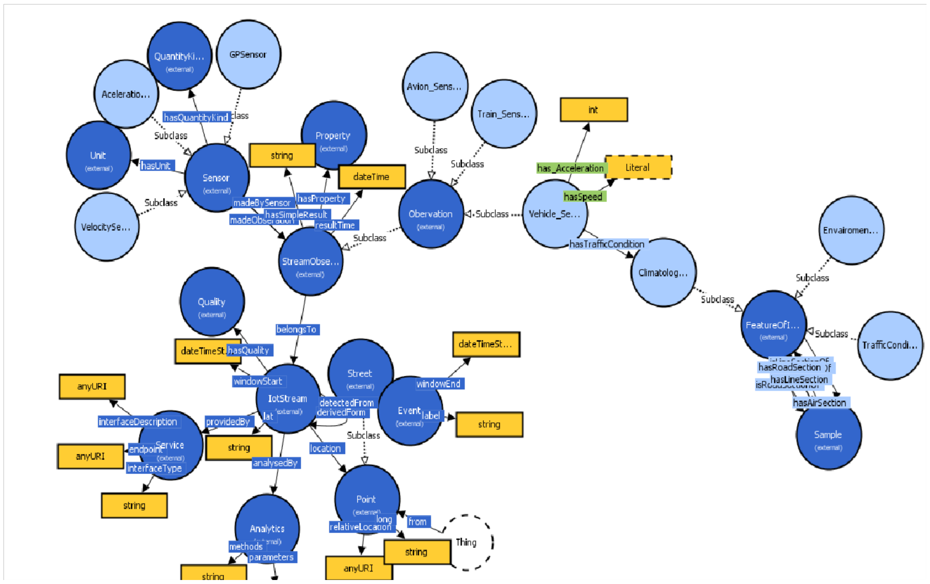

Fig. 1. Trafficstore ontology for semantic annotation

3. Consumer (Client): The client application, responsible for showing produced stream observations and detected events in a suitable way, decoupled from the data sources.
4. Message Broker: The message broker is responsible for receiving and propagating events (new stream observations generated by a producer or derived events generated by some analytic service).
5. Analytics Service 2: The analytics service 2 is responsible for generating events produced by some labeling algorithm, yet to be defined.
6. Registry: The registry is responsible for storing, generating, and exposing (via SPARQL) sensor and stream metadata.

3.3 Annotation System

The annotation process is shown in figure 3. The frequency of the data generated by the sensors is five minutes and is set to represent patterns for each hour. The representation of the hourly pattern, the size of the movement in this case would be 12. For the analysis, the data is divided into windows of 12 data points and LPR is applied [16] in every window. The result is hourly data patterns. Then apply the LPR algorithms [16], the Fuzzy C-means (RI) algorithm [18] is used to apply grouping on the patterns into three different groups. By looking at the centers of the cluster, each group is given a label.

In Analytics services, "M" and "P" correspond to the methods and parameters applied to incoming *StreamObservations*. Seven IotStream Producers publish *StreamObservations* to the broker, and two analytics services subscribe to

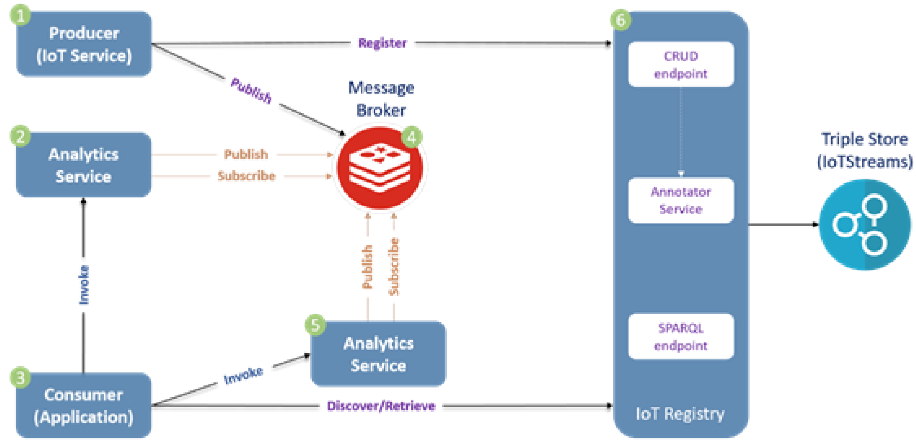

Fig. 2. System Architecture Scheme

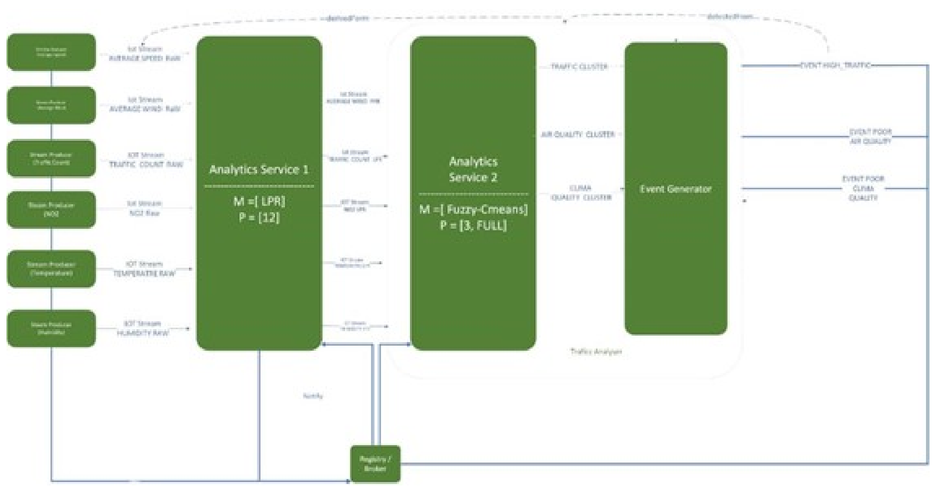

Fig. 3. Annotation system

StreamObservations notifications from the broker. In the case of Analytic Service 1, the *StreamObservations* is received within one hour and the data is processed using LPR [16], and the output is annotated with an IotStream [6][4] instance which was *derivedFrom* and published in the broker. The Analysis Service 2 is then notified and the Fuzzy C-means [18, 12, 19] algorithm is applied in the analyzed flow observations, and then grouped into the predefined groups. The output of the clusters is sent to an Event Generator. Each logged Event has a

[4] See our ontology online

label and is associated with the IotStream instance from which it was detected. This output is then published in the broker for any consumer.

3.4 Data flow

A system that consumes IoT data streams needs to employ some type of data analysis to handle the degree of volume, speed, intermittent, irregularity, and dimensionality. Libraries and frameworks for popular programming languages have allowed the creation of tools that handle data according to its nature and the expected knowledge obtained. Depending on the application, the tools involve forms of pre-processing, machine learning, or correlation.

The result of such techniques can be fed to enrich a semantic knowledge graph. As described in [6], the Reception of Knowledge (RC) web service allows the consumer to use remote IoT data sources with different cascades of methods to study which one works best for them. By exposing a RESTful interface, RC queries data streams from an IoT data stream store using a SPARQL query with a predefined format for the output variables. In turn, the service will generate a new data stream based on the selected methods and their corresponding parameters. The new *StreamObservations* are then annotated and linked to a new IotStream, with the Analytical details used, and then sent back to the Consumer. Figure 4 illustrates the process.

4 Validation

For the validation of the system, we have carried out an experiment with two datasets used in the research: London Air [13] and Open DK [5]. To evaluate the system, the results of the grouping were evaluated from the following indices PC [4], XB [19], PE [3] and OS [18].

We found that the two datasets have the three well-defined classes. An optimal number of bunches is obtained for each index. All indices report the correct cluster number. The Open DK dataset has three clusters. Because two clusters overlap and one cluster is separated from the rest, this is also an acceptable result. The correct cluster numbers and indexes are correctly identified by 3, as reported in table 2.

Table 2. Result of different Indices.

Dataset	Nuñ.C	PC	PE	XB	YOU
London Air	3	3	3	3	3
Open DK	3	2	3	3	3

We carried out two experiments to evaluate the yield of the system. The first one was guided to evaluate the quantity of events of those detected by the humans compared to the events detected by the fuzzy c-means algorithm. The

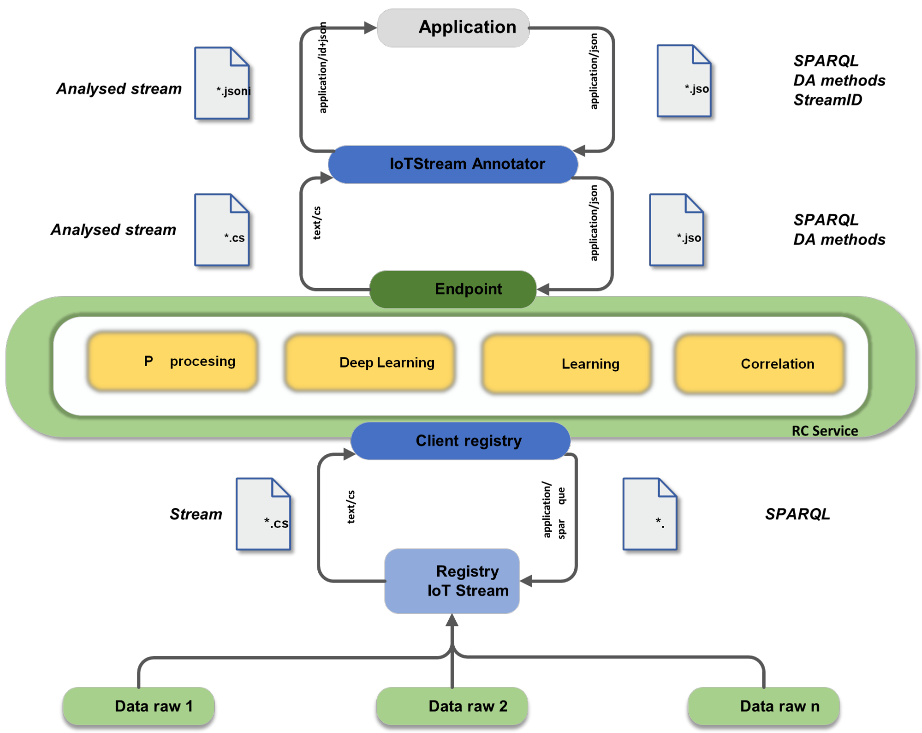

Fig. 4. Data flow

second experiment with the objective to see how many events the system was able to detect in certain time with the algorithm.

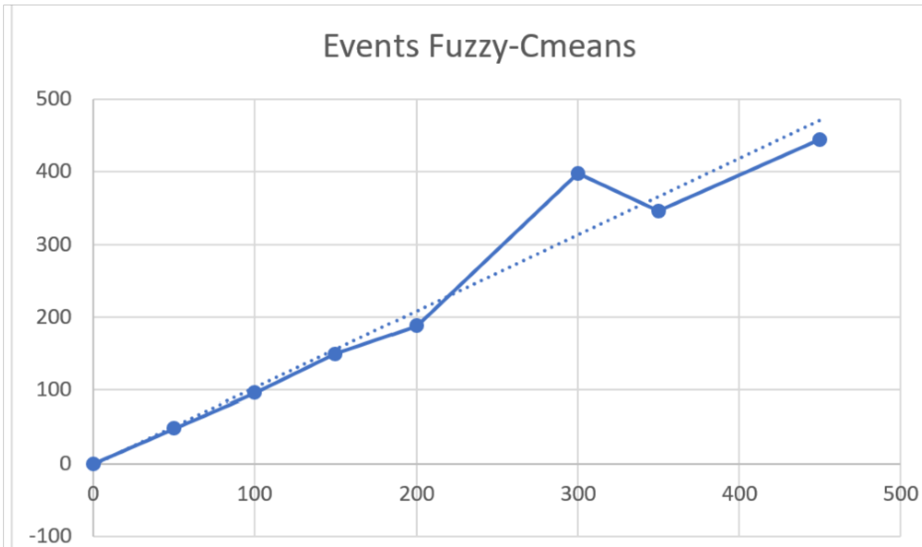

Fig. 5. Humans Events vs Fuzzy C-means.

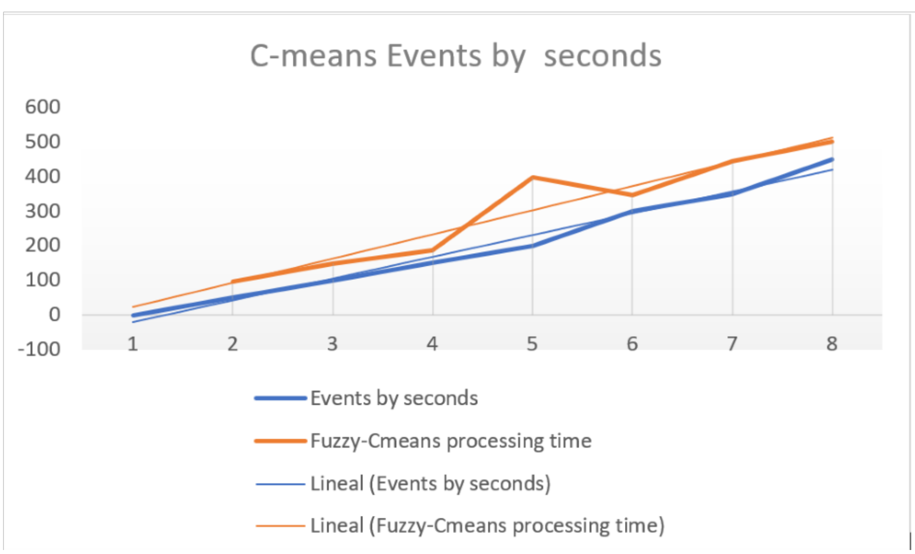

Fig. 6. Events by time.

In the two tests, the results are satisfactory as reported respectively in figures fig. 5 and fig. 6

5 Conclusions

The systems that are currently used for the detection of traffic events are developed with little heterogeneous ontologies. The algorithms used for the detection of events do not allow the evaluation of overlap and sometimes the results that they emit to IoT users have distortions.

The proposed system shows a novelty in the architecture by implementing Open Data schemes and using few components. The evaluation of the results of the use of Fuzzy C-means (FCM) with different indices shows the efficiency of these algorithms in IoT systems when the data even has overlap. It is evident that the use of fuzzy algorithms improves the yield of the IoT system, and it enhances the localization of overlapped data. The ontology developed in this work constitutes a novelty being the first time that a knowledge base is combined with FCM in the transport domain, the ontology improving some pitfalls in previous works.

Although the annotation pattern and the data flow are based on precedent systems. The implementation approach and the novelty in handling the components give validity to precedent investigations in this topic.

Acknowledgments

GA acknowledge grant ANR-19-CE23-0012 from Agence Nationale de la Recherche for project CoSWoT. We thank three anonymous reviewers for their helpful comments.

References

1. Battle, R., Kolas, D.: Enabling the geospatial semantic web with parliament and geosparql. Semantic Web **3**(4), 355–370 (2012)
2. Bermudez-Edo, M., Elsaleh, T., Barnaghi, P., Taylor, K.: Iot-lite: a lightweight semantic model for the internet of things and its use with dynamic semantics. Personal and Ubiquitous Computing **21**(3), 475–487 (2017)
3. Bezdek, J.C.: Cluster validity with fuzzy sets (1973)
4. Bezdek, J.C.: Numerical taxonomy with fuzzy sets. Journal of mathematical biology **1**(1), 57–71 (1974)
5. DK, O.D.: What is open data dk (mar 2015), `https://www.opendata.dk/hvad-er-open-data-dk`
6. Elsaleh, T., Enshaeifar, S., Rezvani, R., Acton, S.T., Janeiko, V., Bermudez-Edo, M.: Iot-stream: A lightweight ontology for internet of things data streams and its use with data analytics and event detection services. Sensors (Basel, Switzerland) **20**(4) (2020). https://doi.org/10.3390/s20040953
7. Gao, F., Ali, M.I., Mileo, A.: Semantic discovery and integration of urban data streams. challenge **7**, 16 (2014)

8. Gómez, S.A., Fillottrani, P.R.: Completitud de los métodos de acceso a datos basado en ontologías: enfoques, propiedades y herramientas. In: XIX Workshop de Investigadores en Ciencias de la Computación (2017)
9. Gorender, S., Silva, Í.: An ontology for a fault tolerant traffic information system. In: 22nd International Congress of Mechanical Engineering (COBEM 2013) (2013)
10. Janowicz, K., Haller, A., Cox, S., Phuoc, D., LefranÃ§ois, M.: Sosa: A lightweight ontology for sensors, observations, samples, and actuators. Journal of Web Semantics **56** (2018). https://doi.org/10.1016/j.websem.2018.06.003
11. Kharlamov, E., Kotidis, Y., Mailis, T., Neuenstadt, C., Nikolaou, C., Özcep, Ö., Svingos, C., Zheleznyakov, D., Brandt, S., Horrocks, I., et al.: Towards analytics aware ontology based access to static and streaming data. In: International Semantic Web Conference. pp. 344–362. Springer (2016)
12. Kim, D.W., Lee, K.H., Lee, D.: On cluster validity index for estimation of the optimal number of fuzzy clusters. Pattern Recognition **37**(10), 2009–2025 (2004)
13. London, I.C.: London air quality network (2021), `https://www.londonair.org.uk/LondonAir/General/about.aspx`
14. Morignot, P., Nashashibi, F.: An ontology-based approach to relax traffic regulation for autonomous vehicle assistance. arXiv preprint arXiv:1212.0768 (2012)
15. Nikolaou, C., Kostylev, E.V., Konstantinidis, G., Kaminski, M., Grau, B.C., Horrocks, I.: The bag semantics of ontology-based data access. arXiv preprint arXiv:1705.07105 (2017)
16. Rezvani, R., Enshaeifar, S., Barnaghi, P.: Lagrangian-based pattern extraction for edge computing in the internet of things. In: 2019 6th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/2019 5th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom). pp. 177–182. IEEE (2019)
17. Tambassi, T.: From a geographical perspective: spatial turn, taxonomies and geo-ontologies. In: The Philosophy of Geo-Ontologies, pp. 27–36. Springer (2018)
18. Wu, K.L., Yang, M.S.: A cluster validity index for fuzzy clustering. pattern recognition letters **26**(9), 1275–1291 (2005)
19. Xie, X.L., Beni, G.: A validity measure for fuzzy clustering. IEEE Transactions on pattern analysis and machine intelligence **13**(8), 841–847 (1991)