AI for Traffic Analytics

Raghava Mutharaju, Freddy Lécué, Jeff Z. Pan, Jiewen Wu, Pascal Hitzler

***Abstract*—Information and communications technology (ICT) is used extensively to better manage the city resources and improve the quality of life of its citizens. ICT spans many departments of the cities, from transportation, water, energy to building management and social-care services. AI techniques are getting more and more attraction from cities to represent and organize information, maintain sustainable networks, predict incidents, optimize distribution, diagnose faults, plan routes and organize their infrastructure. Managing traffic efficiently, among many other domains in cities, is one of the key issues in large cities. In this article we describe the domains of applications which could benefit from AI techniques, along with introducing the necessary background knowledge. Then we focus on traffic applications, which make use of recent AI research in knowledge representation, logic programming, machine learning and reasoning. Specifically we go through the next version of scalable AI driven traffic related application where (1) data from a variety of sources is collected, (2) knowledge about traffic, vehicles, citizens, events is represented and (ii) deductive and inductive reasoning is combined for diagnosing and predicting road traffic congestion. Based on these principles, a real-time, publicly available AI system named STAR-CITY was developed. We discuss the results of deploying STAR-CITY, and its related AI technologies in cities such as Dublin, Bologna, Miami, Rio and the lessons learned. We also discuss the future AI opportunities including scalability issues for large cities.**

***Index Terms*—Artificial intelligence, Knowledge representation, Smart cities, Traffic congestion**

I. Introduction

MORE and more people are moving to the cities in search of better livelihood. The resources and the infrastructure of the cities are unable to keep up with this population growth rate. This leads to several problems such as shortage of water and electricity, increase in pollution and severe traffic congestion, which is one of the major transportation issues in most industrial countries [1]. Traffic congestion leads to massive wastage of time and resources such as fuel. In USA, traffic congestion leads to 5.5 billion hours of delay and 2.9 billion gallons of wasted fuel costing around $121 billion [2]. Apart from such wastage, traffic congestions also lead to road rage and accidents. Three possible ways to reduce traffic congestion [3] are i) improving the road infrastructure, ii) promoting the use of public transport and iii) diagnosing and predicting traffic congestions, which allows city administrators to proactively manage the traffic. Among the three options, the third option is the most cost effective and convenient since it does not involve any change to the existing infrastructure. In this work, we use several AI techniques such as knowledge representation and reasoning, planning and machine learning to predict and diagnose traffic congestions.

There are several existing traffic analysis tools such as US Traffic View[1] [4], French Sytadin[2] and Italian 5T[3]. They support basic analytics, visualization and monitor traffic using dedicated sensors. They cannot handle data coming from heterogeneous sources and do not interpret traffic anomalies. Other systems such as the traffic layer of Google Maps provide real-time traffic conditions but do not take into account the historical data and data from other sources such as weather and city events. Thus the existing systems do not take advantage of the context and the semantics of the data.

Data from several sources provide key insights into the location, cause and intensity of the traffic congestion. User generated content such as tweets, weather conditions, information about city events (music concerts etc) can be used along with the traffic data. Semantic Web technologies such as OWL (Web Ontology Language) [5] and RDF (Resource Description Framework) [6], which are also W3C recommendations, can be used to represent knowledge and integrate data from multiple data sources. These technologies provide structure and meaning to the data as well as enable interlinking, sharing and reuse of the data.

RDF is a framework to describe resources such as documents, people, physical objects, abstract concepts etc. Resources are described in the form of triples, where a triple consists of three parts: subject, predicate and object. For example, we can represent road r_1 is adjacent to road r_2 in the form of a triple as <r_1> <isAdjacentTo> <r_2>.

OWL is more expressive compared to RDF and is used to build ontologies that represent knowledge about things, groups of things and relation between them. It is used to formally encode domain knowledge, i.e., knowledge about some part of the world which is often referred to as the domain of interest. In order to build an ontology, it is important to come up with the vocabulary of the domain, i.e., a set of terms and the relationships between them. These form the axioms in an ontology. The knowledge in an ontology can be categorized into terminological knowledge and assertions. The

Raghava Mutharaju is a research scientist in the Knowledge Discovery Lab of GE Global Research, Niskayuna, NY, USA. Email: raghava.mutharaju@ge.com

Freddy Lécué is a principal scientist and research manager in large scale reasoning systems in Accenture Technology Labs, Dublin, Ireland. He is also a research associate at INRIA, in WIMMICS, Sophia Antipolis, France. Email: freddy.lecue@accenture.com

Jeff Z. Pan is a Reader in the Department of Computing Science at University of Aberdeen, where he is the Deputy Director of Research of the department. Email: jeff.z.pan@abdn.ac.uk

Jiewen Wu is a lead research scientist at Accenture Technology Labs, Dublin, Ireland. Email: jiewen.a.wu@accenture.com

Pascal Hitzler is an endowed NCR Distinguished Professor and Director of Data Science at the Department of Computer Science and Engineering at Wright State University, Dayton, Ohio, USA. Email: pascal.hitzler@wright.edu

[1]https://www.trafficview.org/
[2]http://www.sytadin.fr/
[3]http://www.5t.torino.it/5t/

terminological knowledge or TBox defines the general notions or the conceptualization of the domain whereas the assertional knowledge or ABox defines the concrete notions or facts of the domain. In a database setting, TBox corresponds to the schema and ABox corresponds to the data [7].

Description logics [8], [9] provide the formal underpinnings for OWL. They are fragments of first order logic, with most of them being decidable. They have formal semantics, i.e., a precise specification of the constructs that make up various description logics. This makes them unambiguous and suitable for logical operations. Description logics provide three types of entities: concepts, roles and individual names. Concepts are sets of individuals, roles represent the binary relations between the individuals and individual names represent single individuals in the domain. In first order logic, these three entities correspond to unary predicates, binary predicates and constants. In OWL, concepts and roles are referred to as classes and properties.

The traffic analytics system named STAR-CITY (Semantic Traffic Analytics and Reasoning for CITY) [10] makes use of RDF and description logics to represent the knowledge in the traffic domain and integrate, reason over data from heterogeneous sources [11]. In the rest of the article, we describe the diagnosis and prediction of traffic congestion using STAR-CITY and the lessons learned from deploying STAR-CITY in Dublin (Ireland), Bologna (Italy), Miami (USA) and Rio (Brazil).

Source Type	Data Source	Description	City			
			Dublin (Ireland)	Bologna (Italy)	Miami (USA)	Rio (Brazil)
Traffic Anomaly	Journey travel times across the city	Traffic Department's TRIPS system[a]	CSV format (47 routes, 732 sensors) 0.1 GB per day[b]	✗ (not available)		
	Dublin Bus Dynamics	Vehicle activity (GPS location, line number, delay, stop flag)	✗ (not used)	SIRI: XML format[c] (596 buses, 80KB per update 11GB per day[d])	CSV format (893 buses, 225 KB per update 43 GB per day[e])	CSV format (1, 349 buses, 181 KB per update 14 GB per day[f])
Traffic Diagnosis	Social-Media Related Feeds	Reputable sources of road traffic conditions in Dublin City	"Tweet" format - Accessed through Twitter streaming API[g]			
			Approx. 150 tweets per day[h] (approx. 0.001 GB)	✗ (not available)	Approx. 500 tweets per day[i] (approx. 0.003 GB)	✗ (not available)
	Road Works and Maintenance		PDF format (approx. 0.003 GB per day[j])	XML format (approx. 0.001 GB per day[k])	HTML format (approx. 0.001 GB per day[l])	✗ (not available)
	Social events e.g., music event, political event	Planned events with small attendance	XML format - Accessed once a day through Eventbrite[m]APIs			
			Approx. 85 events per day (0.001 GB)	Approx. 35 events per day (0.001 GB)	Approx. 285 events per day (0.005 GB)	Approx. 232 events per day (0.01 GB)
		Planned events with large attendance	XML format - Accessed once a day through Eventful[m]APIs			
			Approx. 180 events per day (0.05 GB)	Approx. 110 events per day (0.04 GB)	Approx. 425 events per day (0.1 GB)	Approx. 310 events per day (0.08 GB)
	Bus Passenger Loading / Unloading (information related to number of passenger getting in / out)		✗ (not available)	✗ (not available)	CSV format (approx. 0.8 GB per day[e])	CSV format (approx. 0.1 GB per day[e])

[a] Travel-time Reporting Integrated Performance System - http://www.advantechdesign.com.au/trips
[b] http://dublinked.ie/datastore/datasets/dataset-215.php (live)
[c] Service Interface for Real Time Information - http://siri.org.uk
[d] http://82.187.83.50/GoogleServlet/ElaboratedDataPublication (live)
[e] Private Data - No Open data
[f] http://data.rio.rj.gov.br/dataset/gps-de-onibus/resource/cfeb367c-c1c3-4fa7-b742-65c2c99d8d90 (live)
[g] https://sitestream.twitter.com/1.1/site.json?follow=ID
[h] https://twitter.com/LiveDrive - https://twitter.com/aaroadwatch - https://twitter.com/GardaTraffic
[i] https://twitter.com/fl511_southeast
[j] http://www.dublincity.ie/RoadsandTraffic/ScheduledDisruptions/Documents/TrafficNews.pdf
[k] http://82.187.83.50/TMC_DATEX/
[l] http://www.fl511.com/events.aspx
[m] https://www.eventbrite.com/api - http://api.eventful.com

Fig. 1. (Raw) Data Sources used for traffic analytics in Dublin, Bologna, Miami and Rio

II. Semantic Representation and Enrichment of Traffic Data

Traffic on the road can be influenced by a variety of factors such as weather conditions, road works and city events. Accordingly, data from different sources such as sensors, tweets, weather information, city events information etc has to be considered. This can be considered as *Big Data* since the data has all the four important characteristics: volume, velocity, variety and veracity. Figure 1 shows the main attributes of the datasets we considered for traffic analytics.

The next step is to convert the all the heterogeneous data shown in Figure 1 into a homogeneous semantic representation. This representation is useful for comparing and evaluating different contexts e.g., events (and their properties: venue, category, size, types and their subtypes), weather information (highly, moderate, low windy, rainy; good, moderate, bad weather condition). More importantly, semantic representation of data helps in (automatically) designing, learning, applying rules at reasoning time for analysis, diagnosis and prediction components. The static background knowledge and the semantics of the data stream is encoded in an ontology which is in OWL 2 EL profile[4]. $\mathcal{EL}^{++}$ is the description logic underpinning for OWL 2 EL. The selection of the OWL 2 EL profile from among the three OWL 2 profiles has been guided by (i) the expressivity which was required to model semantics of data in our application domain (cf. Figure 1), (ii) the scalability of the underlying basic reasoning mechanisms we needed in our stream context e.g., subsumption in OWL 2 EL is in PTIME [12].

All the data streams in Figure 1 are converted to OWL 2 EL ontology streams using IBM Infosphere Streams [13]. Conversion into streams allows i) easy synchronization and transformation of streams into OWL 2 EL ontology, ii) flexible and scalable composition of stream operations, iii) identification of patterns and rules over different time windows, and iv) possible extension to higher throughput sensors. Depending on the data format, different conversion strategies are used - XSLT for XML, TYPifier [14] for tweets and custom OWL 2 EL mapping for CSV. This is shown in Figure 2.

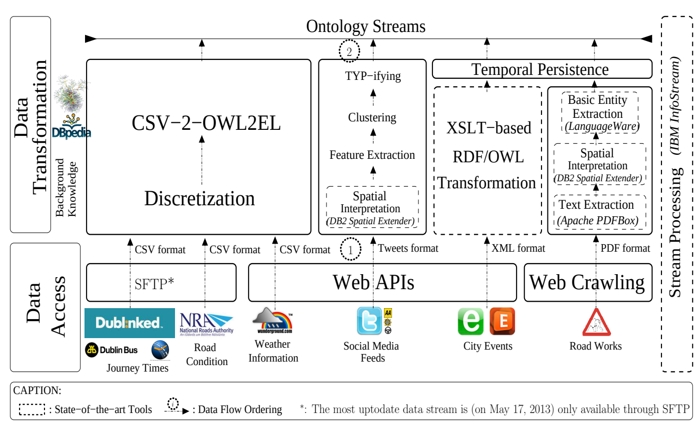

Fig. 2. Semantic Stream Enrichment

[4]https://www.w3.org/TR/owl2-profiles/

III. Diagnosis of Traffic Congestions

Diagnosis task consists of providing a possible explanation for the congestion on a particular road. There can be several reasons for causing or aggravating a traffic congestion. We focus on traffic accidents, road works, weather conditions and social events (e.g., music, political events). Diagnosis of traffic congestion consists of two steps - historic diagnosis computation and real-time diagnosis [15]. This is shown in Figure 3.

All the historic diagnosis information is represented as a deterministic finite state machine. Events, road works and weather conditions are connected to historic traffic congestions along with the probability of those factors indeed causing the congestion. Road intersections and car park locations form the states in the finite state machine. Roads are the transitions in the finite state machine and each road is labeled by its historic diagnosis information.

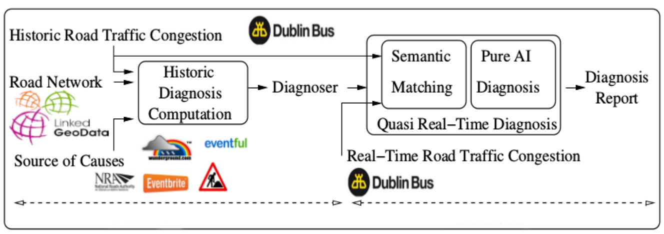

Fig. 3. Overview of the approach to diagnose traffic congestions

After constructing the finite state machine off-line, the next step is to compare the new (current) road condition with the historical condition in real-time and generate a diagnosis report. In existing diagnosis approaches, unless it is an exact match, it is not possible to obtain the diagnosis information. In our approach, we define a matching function that matches the new condition, C_n, which is a description logic concept, with the historical condition, C_h. Note that a condition can be a city event, road work or weather condition which is represented using either existing vocabularies such as DBpedia[5], SKOS[6] or OWL 2 EL ontologies. The matching function gives out the relation between C_n and C_h as output, which could be one of the following.

1) Exact: C_n and C_h are equivalent concepts
2) PlugIn: C_n is a sub-concept of C_h
3) Subsume: C_n is a super-concept of C_h
4) Intersection: The intersection of C_n and C_h is satisfiable

The diagnosis report is constructed using concept abduction between C_n and C_h [16]. The constructed description specifies the under specification in C_h in order to completely satisfy C_n. Computing a diagnosis report is a PTIME problem due to the PTIME complexity of abduction and subsumption in OWL 2 EL.

A crucial step in the diagnosis and prediction of traffic congestions is the classification of ontology streams. Classifying an ontology involves the computation of all the possible sub-concepts for each concept in the ontology. Apart from making implicit sub-concept relationships explicit, classification is also useful for the matching based computation in diagnosis and prediction. Streaming data, which in turn is converted into ontology streams, is considered for the diagnosis and prediction tasks. This would lead to the accumulation of large number of ontologies over a short period of time. Existing reasoners, which are used to classify an ontology, do not scale to large ontologies [17]. A distributed reasoner that can scale with the ontology size is required. DistEL [18] is a distributed and scalable reasoner for the OWL 2 EL profile. An ontology in OWL 2 EL profile can be partitioned based on the different axiom types it supports. Each classification rule of OWL 2 EL (description logic $\mathcal{EL}^{++}$) [19] is applicable to an axiom of one particular type. The partitioned ontology pieces along with the correspond completion rules are distributed across the nodes in the cluster. Each node is dedicated to axioms of at most one particular type and runs the appropriate completion rule on such axioms. This technique improves the data locality and decreases the inter-node communication. A detailed description of other distributed reasoning approaches for OWL 2 EL are described in [20].

IV. Forecasting Traffic Congestions

Predicting or forecasting the anomalies such as traffic congestion involves tracking and correlating the changes (evolution) in the data streams over time [21]. This involves three challenges i) handling the variety and velocity of data (C_1), ii) reasoning on the evolution of multiple data streams (C_2), and iii) scalable and consistent prediction of anomalies (C_3).

The data from different sources (Figure 1) is converted to ontology streams (Figure 2) as discussed earlier. Let $\mathcal{O}_m^n$ represent the journey time and $\mathcal{P}_m^n$ represent the weather information stream from time m to n. $\mathcal{O}_m^n(i)$ is a snapshot of stream $\mathcal{O}_m^n$ at time $i \in [m, n]$. Figure 4 shows the three challenges in predicting journey time using weather information stream. It also captures the weather records and travel conditions on Dame Street at times i, j.

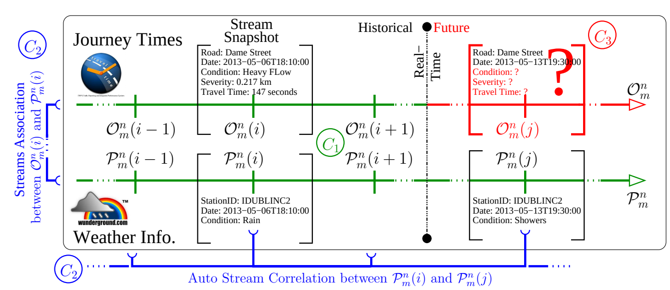

Fig. 4. The challenges (C_1, C_2, C_3) in predicting the journey time $\mathcal{O}_m^n$ using weather stream $\mathcal{P}_m^n$

The second challenge (C_2) is to capture the changes and associate knowledge across the ontology streams. The detection of change along a stream over time enables the computation of knowledge auto-correlation. The semantic similarity between ontology streams is represented by auto-correlation and association aims at deriving rules across streams. These two steps are required to predict the severity of traffic congestions. Prior

[5]http://wiki.dbpedia.org/services-resources/ontology
[6]https://www.w3.org/TR/skos-primer/

to performing the tasks of auto-correlation and knowledge association, it is important to classify the ontology stream. TBox has static knowledge and does not change over time. TBox is generally small and can be classified using the $\mathcal{EL}^{++}$ classification rules [19]. The ontology stream, which is generated from the data stream (Figure 2) consists of ABox axioms. These axioms are internalized into TBox axioms so that the same classification rules from [19] can be applied on them. If existing reasoners (such as CEL[7], ELK[8], Pellet[9] etc) are overwhelmed by the ontology streams, then as discussed earlier, a distributed reasoner such as DistEL [18] can be used.

The auto-correlation between snapshots of an ontology stream is established by comparing the changes in the ABox axioms of the snapshots. The changes can be categorized into three: new, obsolete and invariant. The type of change can have either a positive or a negative influence on the auto-correlation. Invariants have a positive influence on auto-correlation, whereas, new and obsolete changes impact the auto-correlation negatively. Inconsistencies among the snapshots also have a negative correlation. This approach is shown in Figure 5a.

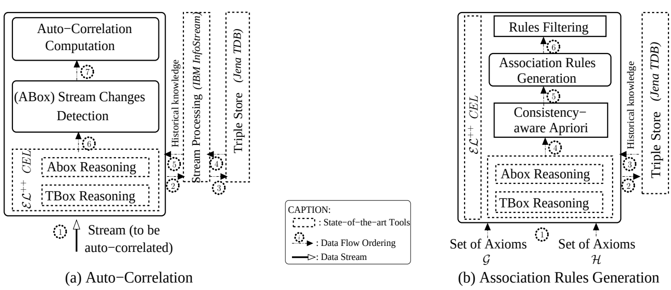

Fig. 5. Auto-correlation among the snapshots of an ontology stream and generation of association rules for prediction

The association rules between the snapshots of a stream are encoded using SWRL[10]. For example, a rule that "*the traffic flow of road r_1 is heavy if r_1 is adjacent to a road r_2 where an accident occurs and the humidity is optimum*" can be represented in SWRL as

$$\text{HeavyTrafficFlow}(s) \leftarrow \text{Road}(r_1) \wedge \text{Road}(r_2) \wedge \text{isAdjacentTo}(r_1, r_2) \wedge \text{hasTravelTimeStatus}(r_1, s) \wedge \text{hasWeatherPhenomenon}(r_1, w) \wedge \text{OptimumHumidity}(w) \wedge \text{hasTrafficPhenomenon}(r_2, a) \wedge \text{RoadTrafficAccident}(a)$$

The generation of association rules is based on a description logic extension of Apriori [22] where subsumption (sub-concept relation) is used to determine association rules. Association is achieved between any ABox elements together with their entailments (e.g., all congested roads, weather, works, incidents, city events, delayed buses). Association is possible only in the case where elements appear in at least one snapshot of the stream. As the number of ABox elements in the stream increases, the number of rules that get generated grows exponentially. Rules are filtered by adapting the definition of *support* (i.e., number of occurrences that support the elements of the rule) and *confidence* (i.e., probability of finding the consequent of the rule in the streams given the antecedents of the rule) for ontology stream. In addition only consistent associations are considered. This approach is shown in Figure 5b. More details on auto-correlation and generation of association rules, including the algorithms, are available in [23].

Although filtering of rules based on support and confidence addresses the scalability concern, it does not however ensure prediction of facts that are consistent (challenge C_3), i.e., facts that do not contradict future knowledge facts. This can be solved by combining auto-correlation with association rule generation. First step is to identify the context (e.g., mild weather, road closure) where the prediction is required, and then perform its auto-correlation with historical contexts. Rules are generated and filtered based on their support, confidence and consistency. A rule is considered as consistent if the consequent of the rule is consistent with the knowledge captured by the exogenous stream [24]. Rules are contextualized and evaluated only against the auto-correlated stream snapshots. This makes the selection of rules knowledge evolution-aware and ensures that rules are applied to contexts where knowledge does not change drastically. This approach of combining auto-correlation with association rule generation is shown in Figure 6.

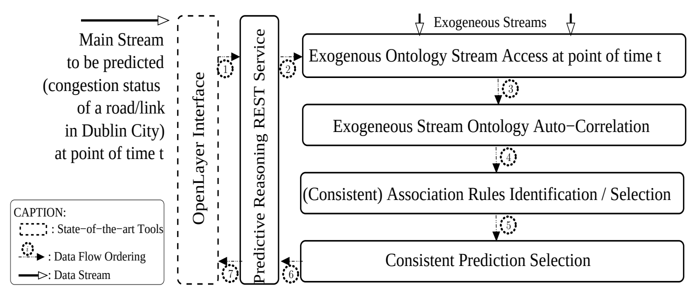

Fig. 6. Auto-correlation is combined with association rule generation for scalable and consistent prediction

V. Lessons Learned

All the features discussed so far, i.e., handling of heterogeneous data, diagnosis and prediction of traffic congestion, have been implemented in a traffic analytics system named STAR-CITY. It makes use of the W3C Semantic Web stack along with other technologies such as i) description logic $\mathcal{EL}^{++}$ based distributed ontology classifier, ii) rule based pattern association, iii) machine learning based entity search, and iv) stream based correlation and inconsistency checking. STAR-CITY was initially deployed in Dublin, Ireland but was later expanded to other cities such as Bologna, Miami and Rio. The challenges and lessons learned in deploying such a system are discussed here.

Heterogeneous streams and semantic expressivity. The format of different data streams (sensors) used in STAR-CITY

[7]https://lat.inf.tu-dresden.de/systems/cel/
[8]https://github.com/liveontologies/elk-reasoner
[9]https://github.com/stardog-union/pellet
[10]https://www.w3.org/Submission/SWRL/

generally remains the same. It is important to pick the right vocabulary and the expressivity to model the data. DBpedia, W3C and NASA ontologies were used to link, integrate and interoperate with all the data sources. However, a custom ontology was developed to model journey time data. Care has to be taken so that terminologies in the various vocabularies used are aligned. This is important in order to achieve on-the-fly integration of all the data sources.

The semantic encoding of city events is in OWL 2 EL profile. This is suitable for us because ontology classification can be decided in polynomial time and hence is scalable. A more expressive profile such as OWL 2 Full or DL could lead to i) more causes getting triggered for road congestion, ii) improving the precision of diagnosis, and iii) improving the scalability and precision of prediction by triggering stronger rules. The downside would be that ontology classification can no longer be done in polynomial time. On the other hand, it would be interesting to check if a profile less expressive than $\mathcal{EL}^{++}$ can be used and still obtain more or less the same precision in diagnosis and prediction.

Scalability of the semantic database. Jena TDB is used to store the semantically enriched data in STAR-CITY. But it could not handle simultaneous updates from multiple streams. So some of the ontology streams had to be delayed in order to accommodate this shortcoming. The B+Trees indexing structure of Jena TDB scales the best in our stream context where large number of updates are performed, i.e., the transaction model is much better handled by this data structure. However there were some scalability issues to handle historical data over more than approximately 110 days. If we do not place any restrictions on the number of days to consider for historical data, then there would be 3,800,000+ events in 458 days. Data gets updated every 20 seconds in this case. In the case of buses, this number is 1000 times larger. Jena TDB cannot handle such large amount of data. Topics such as data, knowledge summarization and stream synchronization needs to be looked into so that the amount of data to be handled by Jena TDB reduces.

Noisy sensor data. Sensors in the real-world exhibit noise. They do not observe the world perfectly due to a number of reasons such as malfunctioning, mis-calibration or network issues. Such noisy data should be detected early so as to avoid unnecessary computations and inaccurate diagnosis, prediction results. In STAR-CITY, some custom filter operators are used to check the validity of the data. These filter operators are defined by analyzing the historical data. For all the data from different sources, the minimum and maximum values are computed. Any record in the data stream that strongly deviates from this interval are removed. If a new data stream is to be considered for traffic analytics, then its historical data needs to be analyzed to determine the appropriate filters. Other mechanisms to filter noisy data should also be looked into.

Temporal reasoning. W3C Time ontology was used to represent the starting data/time and the duration of each snapshot. The temporal similarity between the snapshots of an ontology stream is strictly based on the time intervals. In other words, only the city events and anomalies that match this timer interval are considered. In order to capture more generic temporal aspects such as anomalies during rush hours, bank holidays, weekend, some refinements to the existing ontology are required. Complex features such as temporal intervals could have been used but this could affect the scalability of the application over time. So only basic temporal features were considered. However, more accurate and complext temporal operators could be considered by taking into account the research challenges discussed in [25].

VI. Conclusion

We presented a traffic analytics system named STAR-CITY that can i) handle heterogeneous streaming data from multiple sources, ii) diagnose anomalies such as traffic congestion, and iii) forecast traffic congestions. Heterogeneous data is converted into a homogeneous semantic representation using Semantic Web technologies such as OWL and RDF. In order diagnose traffic congestions, historical data along with other relevant data such as weather information, road works, city events are considered. Concept abduction is used to compare the current event with the historical event and generate a diagnosis report. Forecasting a traffic congestion involves tracking the changes and associating knowledge in the form of rules across the snapshots in an ontology stream. Filtering of rules to avoid rule explosion and consistency in predicting the facts are also discussed. Finally, the lessons learned and the challenges involved in building a scalable traffic analytic system are highlighted.

STAR-CITY supports city managers in understanding the effects of city events, weather conditions and historical data on traffic conditions in order to take corrective actions. It provides valuable insights into real-time traffic conditions making it easier to manage road traffic which in turn helps in efficient urban planning. STAR-CITY has been successfully deployed in some of the major cities such as Dublin (Ireland), Bologna (Italy), Miami (USA) and Rio (Brazil).

References

[1] D. Schrank, B. Eisele, and T. Lomax, "TTI's 2012 urban mobility report," USA, 2012.

[2] R. Arnott and K. Small, "The Economics of Traffic Congestion," *American scientist*, vol. 82, no. 5, pp. 446–455, 1994.

[3] M. Bando, K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama, "Dynamical model of traffic congestion and numerical simulation," *Physical Review E*, vol. 51, pp. 1035–1042, February 1995.

[4] T. Nadeem, S. Dashtinezhad, C. Liao, and L. Iftode, "TrafficView: Traffic Data Dissemination Using Car-to-car Communication," *SIGMOBILE Mobile Computing Communications Review*, vol. 8, no. 3, pp. 6–19, July 2004.

[5] P. Hitzler, M. Krötzsch, B. Parsia, P. F. Patel-Schneider, and S. Rudolph, Eds., *OWL 2 Web Ontology Language: Primer*. W3C Recommendation, 11 December 2012, available from http://www.w3.org/TR/owl2-primer/.

[6] G. Schreiber and Y. Raimond, Eds., *RDF 1.1 Primer*. W3C Recommendation, 24 June 2014, available from https://www.w3.org/TR/rdf11-primer/.

[7] F. Baader, I. Horrocks, and U. Sattler, "Description Logics," in *Handbook of Knowledge Representation*, F. van Harmelen, V. Lifschitz, and B. Porter, Eds. Elsevier, 2008, ch. 3, pp. 135–180.

[8] M. Krötzsch, F. Simančík, and I. Horrocks, "Description Logics," *IEEE Intelligent Systems*, vol. 29, pp. 12–19, 2014.

[9] P. Hitzler, M. Krötzsch, and S. Rudolph, *Foundations of Semantic Web Technologies*. Chapman & Hall/CRC, 2010.

[10] F. Lécué, R. Tucker, S. Tallevi-Diotallevi, R. Nair, Y. Gkoufas, G. Liguori, M. Borioni, A. Rademaker, and L. Barbosa, "Semantic Traffic Diagnosis with STAR-CITY: Architecture and Lessons Learned from Deployment in Dublin, Bologna, Miami and Rio," in *The Semantic Web - ISWC 2014 - 13th International Semantic Web Conference, Riva del Garda, Italy, October 19-23, 2014. Proceedings, Part II*, 2014, pp. 292–307.

[11] S. Kotoulas, V. López, R. Lloyd, M. L. Sbodio, F. Lécué, M. Stephenson, E. M. Daly, V. Bicer, A. Gkoulalas-Divanis, G. D. Lorenzo, A. Schumann, and P. M. Aonghusa, "SPUD - semantic processing of urban data," *Journal of Web Semantics*, vol. 24, pp. 11–17, 2014.

[12] C. Haase and C. Lutz, "Complexity of Subsumption in the [Escr][Lscr] Family of Description Logics: Acyclic and Cyclic TBoxes," in *ECAI 2008 - 18th European Conference on Artificial Intelligence, Patras, Greece, July 21-25, 2008, Proceedings*, 2008, pp. 25–29.

[13] A. Biem, E. Bouillet, H. Feng, A. Ranganathan, A. Riabov, O. Verscheure, H. N. Koutsopoulos, and C. Moran, "IBM Infosphere Streams for Scalable, Real-time, Intelligent Transportation Services," in *Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2010, Indianapolis, Indiana, USA, June 6-10, 2010*, 2010, pp. 1093–1104.

[14] Y. Ma, T. Tran, and V. Bicer, "TYPifier: Inferring the Type Semantics of Structured Data," in *29th IEEE International Conference on Data Engineering, ICDE 2013, Brisbane, Australia, April 8-12, 2013*, 2013, pp. 206–217.

[15] F. Lécué, A. Schumann, and M. L. Sbodio, "Applying Semantic Web Technologies for Diagnosing Road Traffic Congestions," in *The Semantic Web - ISWC 2012 - 11th International Semantic Web Conference, Boston, MA, USA, November 11-15, 2012, Proceedings, Part II*, 2012, pp. 114–130.

[16] T. D. Noia, E. D. Sciascio, F. M. Donini, and M. Mongiello, "Abductive Matchmaking using Description Logics," in *IJCAI-03, Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence, Acapulco, Mexico, August 9-15, 2003*, 2003, pp. 337–342.

[17] K. Dentler, R. Cornet, A. ten Teije, and N. de Keizer, "Comparison of Reasoners for large Ontologies in the OWL 2 EL Profile," *Semantic Web*, vol. 2, no. 2, pp. 71–87, 2011.

[18] R. Mutharaju, P. Hitzler, P. Mateti, and F. Lécué, "Distributed and Scalable OWL EL Reasoning," in *The Semantic Web. Latest Advances and New Domains - 12th European Semantic Web Conference, ESWC 2015, Portoroz, Slovenia, May 31 - June 4, 2015. Proceedings*, 2015, pp. 88–103.

[19] F. Baader, S. Brandt, and C. Lutz, "Pushing the EL Envelope," in *IJCAI-05, Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence, Edinburgh, Scotland, UK, July 30 - August 5, 2005*, 2005, pp. 364–369.

[20] R. Mutharaju, "Distributed Rule-Based Ontology Reasoning," Ph.D. dissertation, Wright State University, 2016.

[21] F. Lécué, R. Tucker, V. Bicer, P. Tommasi, S. Tallevi-Diotallevi, and M. L. Sbodio, "Predicting Severity of Road Traffic Congestion using Semantic Web Technologies." in *Proceedings of the 11th Extended Semantic Web Conference (ESWC2014), Anissaras, Crete, Greece, May 25–May 29, 2014*, 2014.

[22] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules in Large Databases," in *Proceedings of the 20th International Conference on Very Large Data Bases*, ser. VLDB '94, San Francisco, CA, USA, 1994, pp. 487–499.

[23] F. Lécué and J. Z. Pan, "Predicting Knowledge in an Ontology Stream," in *IJCAI 2013, Proceedings of the 23rd International Joint Conference on Artificial Intelligence, Beijing, China, August 3-9, 2013*, 2013, pp. 2662–2669.

[24] J. Wu and F. Lécué, "Towards Consistency Checking over Evolving Ontologies," in *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, CIKM 2014, Shanghai, China, November 3-7, 2014*, 2014, pp. 909–918.

[25] C. Lutz, "Combining interval-based temporal reasoning with general TBoxes," *Artificial Intelligence*, vol. 152, no. 2, pp. 235–274, 2004.