

Association Rule Mining with Context Ontologies: An Application to Mobile Sensing of Water Quality

Eliot Bytyçi^{1(✉)}, Lule Ahmedi¹, and Arianit Kurti^{2,3}

¹ University of Prishtina, Prishtina, Kosovo
{eliot.bytyci, lule.ahmedi}@uni-pr.edu

² Linnaeus University, Växjö, Sweden
arianit.kurti@lnu.se

³ Interactive Institute Swedish ICT, Norrköping, Sweden
arianit.kurti@tii.se

Abstract. Internet of Things (IoT) applications by means of wireless sensor networks (WSN) produce large amounts of raw data. These data might formally be defined by following a semantic IoT model that covers data, meta-data, as well as their relations, or might simply be stored in a database without any formal specification. In both cases, using association rules as a data mining technique may result into inferring interesting relations between data and/or metadata. In this paper we argue that the context has not been used extensively for added value to the mining process. Therefore, we propose a different approach when it comes to association rule mining by enriching it with a context-aware ontology. The approach is demonstrated by hand of an application to WSNs for water quality monitoring. Initially, new ontology, its concepts and relationships are introduced to model water quality monitoring through mobile sensors. Consequently, the ontology is populated with quality data generated by sensors, and enriched afterwards with context. Finally, the evaluation results of our approach of including context ontology in the mining process are promising: new association rules have been derived, providing thus new knowledge not inferable when applying association rule mining simply over raw data.

Keywords: Association rules · Ontology · Context · Wireless sensor networks · Internet of Things

1 Introduction

In the era of the Internet of Things, wireless sensor networks (WSN) play a major part, due to their affordability. The problem, however arises in collection, exchange and analysis of the data produced. These data, referred to as raw data initially, are collected directly from the source and streamed towards data repositories for further handling. However, this leads to problems we usually encounter with raw data in terms of data quality, heterogeneity and the space needed for their preservation.

Nevertheless, those problems have been mitigated using different scenarios, in order to minimize the effect of the challenges obtained. But, some new challenges have come forward in the last few years. One of them is also the semantic integration of data. Meaning of a concept and their description is basis of the semantic study. That challenge might be resolved by using the ontologies for description of concepts and their relations [7]. Furthermore, new relations can be drawn from the ontology representation of initial concepts, thus improving not only the quality of the data but checking also possible inconsistencies occurred during data integration.

Lastly, data needs to be analyzed in order to discover useful information, which is the reason why those data have been initially gathered and processed. The analysis can be descriptive, using data to summarize important components, and predictive, using data to predict further relations and discover new knowledge. Knowledge discovery in databases (KDD) is the field concerned with developing techniques and methods for making sense of data [8]. One of the stages of the process is data mining, which besides different tasks involves also association rule mining. Association rule mining represents a powerful method for discovering relations between data [9]. By using it jointly with ontology, we consider that further interesting relations can be drawn.

The relations drawn, can vary from one situation to another. Furthermore, in order to entirely understand them, one needs to know the circumstances that create that situation. That represents the context upon which rules are formed. In [1], authors showed that context could provide accuracy and efficacy to data mining outcomes used in medical applications.

In our case, we tend to use association rule mining with context ontologies in surface water quality monitoring. The monitoring is performed through mobile sensing devices, which measure several parameters and forward them to a repository. This component is part of a bigger system, which involves also static monitoring stations for water quality monitoring. Such application can be further extended for usage in other domains.

The paper is organized as follows: in the next section we provide insight on related work, while in Sect. 3 we describe data preprocessing process. In Sect. 4, our ontology modeling for mobile sensing of water quality is described, with the context inference module included and in Sect. 5, the results from association rule mining on data and on context-aware ontology data are presented. Conclusions, challenges and future contribution are covered in Sect. 6.

2 Related Work

The authors of [1] have introduced a framework of representing context in ontology, firstly captured during data mining process and then adapting it accordingly. They have used a classification tree to predict accurately the patient's heart attack risk. In the end, their framework showed that use of the context factor had increased effectiveness in data mining.

Authors in [13] have addressed the challenge of mining knowledge encoded in domain ontologies. They have demonstrated the usefulness of their approach by mining biological data, showing major improvements and advantages.

In [14], authors have mined with association rules over RDF and OWL data repositories. The appropriate transactions were derived from ontology through schema

knowledge for further mining through association rules algorithms. Their initial experiments have proved usefulness and efficiency of the approach.

The concept of “mining configurations” has been introduced by authors in [15], allowing mining of RDF data at different levels. Among configurations is the one describing relation among the subjects and the objects in the RDF triples through basket analysis. Authors at the same time call for further research in the field of association rule mining over RDF by combining configurations and different use cases.

The discussed related approaches are characterized by the combination of ontologies and data mining, but none of them have used association rules, and generated context and ontology to then further advance with data mining of the Semantic Web. That has served us as a motivation for further work that related these three concepts: association rule mining, ontologies, and context.

3 Background and Hypothesis

Internet of Things relies on sensors to monitor the environment. Sensors produce data that should be further processed and analyzed in order to infer new knowledge. New knowledge should help us on differing between a usual or unusual process happening in the environment. That can be used to respond to the environment with possible actuators.

3.1 Problem Definition – Water Quality Case Study

Let us consider an example where sensors are used to measure values of water quality parameters in a river. The measurements are performed for several parameters such as pH, water temperature, dissolved oxygen and conductivity. In [2], it is acknowledged that during the night the values of dissolved oxygen will fall sharply, mainly due to the process of photosynthesis that occurs only during the day. Another parameter related to that is temperature, which during the night is lower due to the deprivation of sun. Related to both parameters, an elevated turbidity can increase the water temperature and will lower the dissolved oxygen (DO), imitating the process of photosynthesis as a regular process that can happen during the night. Furthermore, as presented in [3], a direct variation exists as a correlation between pH and temperature. Therefore, hypothetically, if the dissolved oxygen falls sharply and temperature is lower as well, resulting in lower pH, we can observe this as a regular process if it occurs during the night. But, if the same process is happening during the day, we can treat it as something out of ordinary as a possible ongoing pollution. Hence, it depends if it is day or night, in order to make the substantial difference of the process.

3.2 The InWaterSense Project

InWaterSense project [5], which contributed with a wireless sensor network deployed in river Sitnica, has a static component consisting of several sensors, in order to measure the water quality parameters. In addition to the static part, the deployed system consists

also of a mobile component, with the aim to discover other possible polluted water locations for the deployment of the static system in the future. The mobile component measured 4 parameters: pH, temperature, dissolved oxygen and conductivity. Using an open source wireless platform, with sensors attached to it, measurement data, herein after raw data, was sent in real time to a specific remote server for storage to a database, with timestamp data attached.

After several measurements, data were manually analyzed, where several anomalies were observed, which back up the need for simulated data. For example, in several cases, the temperature was measured below 0 or in some cases pH values were -1 . After considerable measurements, it was concluded that those outlier data could have been result of several conditions:

- sensors have made measurements before entering to water: this due to the fact that all the measurements were done from the bridge (sensors were lowered to the water) due to impossible approach to the specific locations from nearby river.
- sensors were not calibrated: a periodic calibration of the sensors is mandatory, or
- sensor damage: at least in one case sensor was damaged due to fast water streams.

Another prevailing factor that determined usage of simulated data was the amount of data stored, less than 1000 records. Therefore, with a generator created as part of the web portal in project [16], data was simulated in large amounts. Besides that, the generator was used carefully in order to control values of data, so to back up claims by the water experts on their correlation. Furthermore, backing our initial hypothesis, that due to photosynthesis during the night the value of DO will be lower, we have created a constraint in generation of DO to be lower or equal to DO during day hours. If there is a rapid fall of DO during the day, then the reason should be searched with turbidity or direct pollution, according to the experts. Using the data generator, more than 100000 records were generated comprising of timestamp and sensor values.

After the generation of the data and before starting the process of data evaluation, a process of data pre-processing was concluded. The first step of preprocessing the division of the timestamp into several parts including: year, month, date, hour and minutes for the purposes of finding the day or night interval. For experimental purposes, day was described from 6 h in the morning until 18 h in the evening and in contrary the night was described as from 18 h in the evening until 6 h in the morning. After that, data were discretized or divided into several bins, each maintaining data with similar characteristics. A number of three bins were chosen while dividing parameters: a bin that holds parameters with lower values, those with medium values and lastly the ones with higher values. The automatic process of bin division was performed with help of a support tool - WEKA [11]. After that, we have removed some data that were seen unnecessary such as year, day, minutes, since they were repetitive and therefore not significant for the process. In the end, a final set of data was obtained for further processing and analyzing.

Similar preprocessing was conducted over Ontology, were additionally, in the beginning of the process, a data cleaning and preprocessing was conducted. That meant removing some of the relations between ontology concepts and concepts descriptions, due to the purpose of the experiment purely on data. Affirmatively, the process of

ontology population was performed before hand, a process that will be described in the next section.

The final obtained set, both data and ontology data, was ready for the experiment of data mining. Data mining techniques that have been widely used to find patterns of mining data are Apriori and FPGrowth. Association rule mining may return interesting relations between values, in our case values obtained through water quality measurement performed by sensors, thus identifying new rules that describe correlations between data and have a certain percentage of trust [10]. Apriori has been used in databases containing transactions in order to find correlations between the items on such transaction. The values obtained from water quality measurements through sensors, can be viewed as such transactions. A transaction is comprised of timestamp and parameter values. Thus, using association rule mining, one can explore correlations between values. That would result in generation of new rules between sensors obtained values. Such rules would acknowledge a possible failure in the system or a possible dangerous situation, in regards to water quality monitoring and pollution. Besides that, backing our initial assumption, one could claim that by using Ontology, we can extract even more rules, which would result in new knowledge being revealed. The input from the ontologies would be in the context provided and therefore enhancing the knowledge base.

Therefore, in this paper we aim that by using simulated data, that imitate increase or decrease of specific sensor values during night, in conjunction with association rule mining and ontology, we would be able to derive new rules. To achieve that, ontology should be created and populated with generated data, before providing the necessary context for the inference of new knowledge.

4 Context Ontology

In order to formally define the entities of the mobile water quality monitoring sensor system, a lightweight ontology is introduced. The ontology is populated with generated sensor data. As known from the literature [7], ontology describes the overall agents involved in the system and their relations. Besides them, i.e., the mobile sensing component and the water quality related data it generates through measurements, the context is also covered by our newly introduced ontology. That motivated in our example domain by known facts, e.g., that due to photosynthesis during night, certain observed water quality parameters take different values when compared to their values during day. Thus, the day/night context enables modeling these domain specific context behaviors and their implications, as will be made explicit in the examples to follow.

Figure 1 depicts the proposed lightweight mobile context-aware ontology named LMINWS¹. Whereas authors in [4] modeled an ontology (InWaterSense) that covers an arbitrary wireless sensor network for water quality monitoring, this lightweight ontology aims to cover modeling the rather more rich-in-context but simple mobile portable sensors of a wireless sensor network. The classes of our lightweight ontology are depicted in grey color, in Fig. 1, while the ones in white represent imports from other

¹ <http://inwatersense.uni-pr.edu/ontologies/LMINWS.owl>.

ontologies. The Time² ontology, its class `DateTimeDescription` has been extended with addition of two new subclasses in order to represent the day and night context. The `DayDescription` subclass expresses the time between 6 o'clock in the morning and 18 o'clock. The rest of time, from 18 o'clock until 6 o'clock in the morning is modeled as `NightDescription`. Another additional core class in our ontology is the `MobileEquipment`, representing only the mobile sensing part of the WSN system for water quality monitoring, and specializes as such the `InWaterSense` ontology introduced in [4]. An important concept introduced, which is aims to serve for future work, is the `Activity` class with its two subclasses `CalibrationActivity` and `MeasuringActivity`. It helps on context implementation when related to the user who performs the activity: an `Engineer` or a `Technician`. Both later concepts are introduced in the ontology as well belonging to the class `Person` of the FOAF³ ontology, in order to determine by whom exactly the given activity is conducted, and whether data can be reliable or not. Another concepts used is `Place`⁴, with two other subclasses introduced `InDoor` and `OutDoor`.

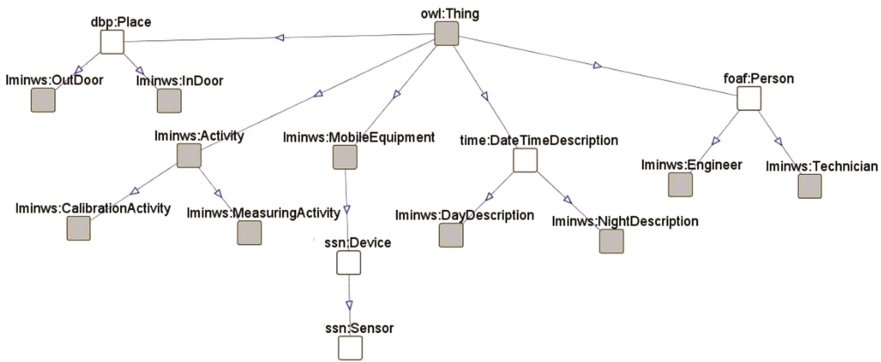


Fig. 1. Lightweight mobile ontology (LMINWS)

4.1 Populating the Ontology

Once sensors generate the data, the modeled ontology is populated right away. Since the tool used [12] for populating the ontology through mapping required the specific format of data representation, before populating it we have converted data to the requested specific format. Then a mapping file, partially presented in Fig. 2, was created in order to convert database data into an RDF/XML format. Subsequently data was added to the existing ontology as a repository.

² <https://www.w3.org/TR/owl-time/>.

³ <http://xmlns.com/foaf/spec/>.

⁴ <http://dbpedia.org/ontology/>.

```

@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix d2rq: <http://www.wiwiss.fu-berlin.de/suhl/bizer/D2RQ/0.1#> .
@prefix jdbc: <http://d2rq.org/terms/jdbc/> .

map:database a d2rq:Database;
  d2rq:jdbcDriver "com.mysql.jdbc.Driver";
  d2rq:jdbcDSN "jdbc:mysql://localhost/sensordata";
  d2rq:username "root";
  jdbc:autoReconnect "true";
  jdbc:zeroDateTimeBehavior "convertToNull";
  .

# Table data
map:data a d2rq:ClassMap;
  d2rq:dataStorage map:database;
  d2rq:uriPattern "data/@@data.Id@";
  d2rq:class vocab:data;
  d2rq:classDefinitionLabel "data";
  .

map:data__label a d2rq:PropertyBridge;
  d2rq:belongsToClassMap map:data;
  d2rq:property rdfs:label;
  d2rq:pattern "data #@@data.Id@";

```

Fig. 2. Partially described data mapping file

During the conversion process, several specific issues were encountered which needed manual intervention, especially when dealing with prefixes.

4.2 Context Inference

Once the ontology is populated with data, then by means of an ontology reasoner, new context-dependent data may get inferred from existing ontology data.

Let us consider again the running day/night context example, in `DateTimeDescription` class in the ontology (cf. Fig. 1), two subclasses are introduced: `DayDescription` and `NightDescription`. Initially, date and time data are assigned instances of `DateTimeDescription`. After applying the hour constraint, those instances are inferred as instances of either `DayDescription` subclass or the `NightDescription` subclass.

The constraint on `DayDescription`, subclass of `DateTimeDescription`, using the data property `onProperty: hour`, was introduced as follows:

$$\textit{hour some xsd:int}[\geq="0"^^\textit{xsd:int}, \leq="17"^^\textit{xsd:int}]$$

It infers only instances of `DateTimeDescription` that belong to the day description, meaning only measurements performed during day hours.

A similar constraint is defined for `NightDescription`, by simply putting the negation over `DayDescription` instances as follows:

$$\text{not (hour only xsd:int[>="6"^^xsd:int, <="17"^^xsd:int])}$$

In the end, the lightweight ontology triplets were derived, describing ontology concepts and their relations. A snippet of these triplets may be seen in Fig. 3, where concepts such as e.g. MobileComponent can be seen standing in relation with NightDescription and the corresponding values of the water quality measurement sensors.

@type/1	@type/2	#hasConductivity	#hasDissolvedOxygen
#MobileEquipment	#NightDescription	\{(3383.333333-inf)\}	\{(100-200)\}
#MobileEquipment	#NightDescription	\{(3383.333333-inf)\}	\{(100-200)\}
#MobileEquipment	#NightDescription	\{(1766.666667-3383.333333)\}	\{-inf-100\}
#MobileEquipment	#NightDescription	\{(1766.666667-3383.333333)\}	\{-inf-100\}
#MobileEquipment	#NightDescription	\{-inf-1766.666667\}	\{-inf-100\}
#MobileEquipment	#NightDescription	\{(1766.666667-3383.333333)\}	\{-inf-100\}

Fig. 3. Portion of triplets generated from ontology

5 Association Rule Mining with Context Ontology

In our deployed wireless sensor network for water quality monitoring, its static and mobile sensors generate data, which are then sent to a remote server, as presented in Fig. 4. Those transactions include values of the sensor measurements on the water quality parameters and the timestamp when the measurement occurred. From the previous Section, the context ontology part of the architecture was explained, preceded by the description of the problem in Sect. 3. From the ontology, data were transformed into JSON format. That because the requested format for the tool used - WEKA [11] was CSV or Arff. After that, on preprocessed data, algorithms were used, which resulted in gaining new knowledge.

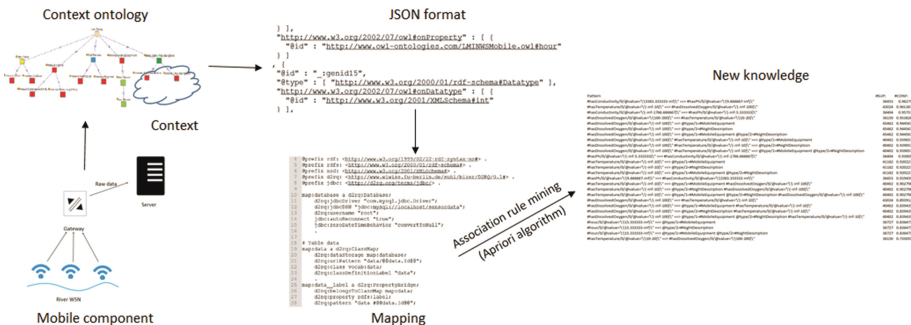


Fig. 4. The architecture of the system

Association rule mining is based on the market basket analysis, which analyses transaction repository [9]. In the repository, there exist sets of items, which once may

describe by X and Y . Therefore, the association rule expressed as $X \Rightarrow Y$, denotes that in the transaction database, a number of transactions contain X , with a certain probability that they will contain also Y .

In [6], authors have presented top ten most used algorithms in data mining. Amongst them is the Apriori algorithm [10], which is used for finding frequent subsets (item sets) from a transaction dataset and derive association rules. When there are no other subsets, the algorithm stops. It should be noted that one can limit the *support* threshold of the algorithm so that it generates only transactions that have a specific number of appearances. Besides that, also the *confidence* can be limited by user in order to find those transactions that contain both items of the rule. Best rules are those with the higher support and confidence.

An improvement over Apriori is the so-called FP-growth (frequent pattern growth) method that succeeds in eliminating candidate generation [6] used in Apriori.

We have conducted tests with both algorithms Apriori and FP Growth over our data, but have always obtained same results. Therefore, only results obtained when applying one of the algorithms, i.e. Apriori, will be presented next.

Data preparation. The sensor measurement data are preprocessed for the mining process by using several unsupervised methods such as normalization and discretization of data. Besides that, since we need to know the timing of data measured, we have split the timestamp into smaller pieces. One of those smaller pieces is the hour when the measurement occurred, which actually provides the context on whether the measurement happened during day or night. That may help on realizing which association rules generated are out of ordinary. Moreover, in order to find significant rules, data have first been discretized – a filter that allows distribution of data in separate bins. As previously mentioned, we have distributed our data into 3 bins proportionally depending on the values, with the help of a tool - WEKA [11]. For example, the temperature is divided into low temperature, mid temperature and high temperature.

Running example. The architecture presented in Fig. 4, depict the running example of the system in water quality monitoring. It explains the process and the solution according to our approach of a specific domain problem.

Results of using Apriori algorithm over sensor data enriched with the LMINWS context ontology are shown in Fig. 5.

Observing Fig. 5, one can see there exists a relation between temperature and day or night context. The association rule that during night, the temperature is lower, has bigger support and higher confidence. Further similar association rules exist, amongst others related to dissolved oxygen, which has lower values during night, as well.

Results and discussion. We consider that our approach of aiding association rule mining with context ontologies is more successful on providing inferred knowledge than existing approaches in the literature which proclaim using association rule mining over simply raw data [14, 16]. To back up that claim, we have performed an experiment only on raw data, i.e., without the context ontology, and compared the results to the previous example.

PATTERN	SUP	CONF
#hasConductivity/0/@value='\{3383.333333-inf}' ==> #hasPh/0/@value='\{9.666667-inf}'	36453	0.96271
#hasTemperature/0/@value='\{-inf-10}' ==> #hasDissolvedOxygen/0/@value='\{-inf-100}'	43024	0.961387
#hasConductivity/0/@value='\{-inf-1766.666667}' ==> #hasPh/0/@value='\{-inf-5.333333}'	36404	0.95737
#hasDissolvedOxygen/0/@value='\{100-200}' ==> #hasTemperature/0/@value='\{10-20}'	36150	0.953826
#hasDissolvedOxygen/0/@value='\{-inf-100}' ==> @type/1=#MobileEquipment	45462	0.944567
#hasDissolvedOxygen/0/@value='\{-inf-100}' ==> @type/2=#NightDescription	45462	0.944567
#hasDissolvedOxygen/0/@value='\{-inf-100}' ==> @type/1=#MobileEquipment @type/2=#NightDescription	45462	0.944567
#hasDissolvedOxygen/0/@value='\{-inf-100}' #hasTemperature/0/@value='\{-inf-10}' ==> @type/1=#MobileEquipment	40402	0.939057
#hasDissolvedOxygen/0/@value='\{-inf-100}' #hasTemperature/0/@value='\{-inf-10}' ==> @type/2=#NightDescription	40402	0.939057
#hasDissolvedOxygen/0/@value='\{-inf-100}' #hasTemperature/0/@value='\{-inf-10}' ==> @type/1=#MobileEquipment @type/2=#NightDescription	40402	0.939057
#hasPh/0/@value='\{-inf-5.333333}' ==> #hasConductivity/0/@value='\{-inf-1766.666667}'	36404	0.93005
#hasTemperature/0/@value='\{-inf-10}' ==> @type/1=#MobileEquipment	41182	0.920227
#hasTemperature/0/@value='\{-inf-10}' ==> @type/2=#NightDescription	41182	0.920227
#hasTemperature/0/@value='\{-inf-10}' ==> @type/1=#MobileEquipment @type/2=#NightDescription	41182	0.920227
#hasPh/0/@value='\{9.666667-inf}' ==> #hasConductivity/0/@value='\{3383.333333-inf}'	36453	0.919439
#hasTemperature/0/@value='\{-inf-10}' ==> @type/1=#MobileEquipment #hasDissolvedOxygen/0/@value='\{-inf-100}'	40402	0.902798
#hasTemperature/0/@value='\{-inf-10}' ==> @type/2=#NightDescription #hasDissolvedOxygen/0/@value='\{-inf-100}'	40402	0.902798
#hasTemperature/0/@value='\{-inf-10}' ==> @type/1=#MobileEquipment @type/2=#NightDescription #hasDissolvedOxygen/0/@value='\{-inf-100}'	40402	0.902798
#hasDissolvedOxygen/0/@value='\{-inf-100}' ==> #hasTemperature/0/@value='\{-inf-10}'	43024	0.899112
#hasDissolvedOxygen/0/@value='\{-inf-100}' ==> @type/1=#MobileEquipment #hasTemperature/0/@value='\{-inf-10}'	40402	0.839435
#hasDissolvedOxygen/0/@value='\{-inf-100}' ==> @type/2=#NightDescription #hasTemperature/0/@value='\{-inf-10}'	40402	0.839435
#hasDissolvedOxygen/0/@value='\{-inf-100}' ==> @type/1=#MobileEquipment @type/2=#NightDescription #hasTemperature/0/@value='\{-inf-10}'	40402	0.839435
#hour/0/@value='\{15.333333-inf}' ==> @type/1=#MobileEquipment	36727	0.836473
#hour/0/@value='\{15.333333-inf}' ==> @type/2=#NightDescription	36727	0.836473
#hour/0/@value='\{15.333333-inf}' ==> @type/1=#MobileEquipment @type/2=#NightDescription	36727	0.836473
#hasTemperature/0/@value='\{10-20}' ==> #hasDissolvedOxygen/0/@value='\{100-200}'	36150	0.735055

Fig. 5. Results of association rule mining over sensor data with LMINWS context ontology

Thus, the Apriori algorithm over raw data has been applied to find all possible rules. Results are presented in Fig. 6.

PATTERN	SUP	CONF
Conductivity='\{3383.333333-inf}' ==> pH='\{9.666667-inf}'	36479	0.96271
Temperature='\{-inf-10}' ==> DO='\{-inf-100}'	43048	0.961386
Conductivity='\{-inf-1766.666667}' ==> pH='\{-inf-5.333333}'	36430	0.957374
DO='\{100-200}' ==> Temperature='\{10-20}'	36168	0.953823
pH='\{-inf-5.333333}' ==> Conductivity='\{-inf-1766.666667}'	36430	0.930001
pH='\{9.666667-inf}' ==> Conductivity='\{3383.333333-inf}'	36479	0.919399
DO='\{-inf-100}' ==> Temperature='\{-inf-10}'	43048	0.893891
Temperature='\{10-20}' ==> Day='\{12.666667-13.333333}'	38362	0.779541
Temperature='\{10-20}' ==> DO='\{100-200}'	36168	0.734958
Day='\{12.666667-13.333333}' ==> Temperature='\{10-20}'	38362	0.445174

Fig. 6. Apriori over raw data

At the first glance, we observe there are a set of rules derived with Apriori for the same running example on both cases, with and without the context ontology. As has been required, only the rules that have the support more than 30 % are provided on the result set. That was the same condition that we have put for both experimental cases. We observe that in both cases, obtained rules that relate parameters such as conductivity and pH, have high confidence. But, in the experiment with the LMINWS context ontology, we have derived new rules not derived when running the same experiment but without context ontology e.g. the rules that involve context. Therefore, context modeled through the ontology and considered while mining obviously makes the difference. Therefore, as expected in the beginning of running example, we have obtained results that relate certain parameters with the context, i.e., day or night context. This may assist

experts in concluding whether certain inferred relations are confident or happening as a result of the natural process, such as photosynthesis. This in addition to the other fewer in number obtained rules when applied over raw data, which describe rather aid in inferring relations between parameters and are not related to the context.

6 Conclusion, Challenges and Future Work

In this paper we have presented an approach of using association rule mining over mobile component of wireless sensor network data, with context-aware ontology. Initially, the concepts and relationships to model a mobile water quality measurement sensing device through an ontology have been described. The ontology is also enriched by contextual concepts and restrictions. Following that, we have populated the ontology with the sensor measurement data. Only then, using association rule techniques, it is proved that the achieved results are richer, compared to the results obtained when association rule mining techniques are used over raw data and without context ontology. An increasing number of rules are obtained, in case of association rule mining with context ontology, i.e., with the LMINWS lightweight context ontology in our example domain. That was verified by the same experiment performed but over raw data.

We haven't been able to find a specific approach similar to ours on finding rules related to context by hand of a context-aware ontology. Furthermore, there is to the best of our knowledge no such a study in the domain of water quality monitoring domain and with wireless mobile sensors. Using the same approach, we aim to extend the experiments to other domains such as health in the future. Moreover, we will check with other data mining techniques to divide data into bins, since we believe that it could yield even better results.

References

1. Singh, S., Vajirkar, P., Lee, Y.: Context-based data mining using ontologies. In: Song, I.-Y., Liddle, S.W., Ling, T.-W., Scheuermann, P. (eds.) ER 2003. LNCS, vol. 2813, pp. 405–418. Springer, Heidelberg (2003). doi:[10.1007/978-3-540-39648-2_32](https://doi.org/10.1007/978-3-540-39648-2_32)
2. “Water Quality” (2016). <http://www.grc.nasa.gov/WWW/k-12/fenlewis/Waterquality.html>. Accessed 24 June 2016
3. (2016) <http://depts.alverno.edu/nsmt/archive/SagatClarkNathavong.htm>. Accessed 24 June 2016
4. Jajaga, E., Ahmedi, L., Ahmedi, F.: An expert system for water quality monitoring based on ontology. In: Garoufallou, E., Hartley, R.J., Gaitanou, P. (eds.) MTSR 2015. CCIS, vol. 544, pp. 89–100. Springer, Heidelberg (2015). doi:[10.1007/978-3-319-24129-6_8](https://doi.org/10.1007/978-3-319-24129-6_8)
5. Ahmedi, L., Jajaga, E., Ahmedi, F.: An ontology framework for water quality management. In: Proceedings of the 6th International Conference on Semantic Sensor Networks, vol. 1063, pp. 35–50. CEUR-WS (2013)
6. Wu, X., Kumar, V., Quinlan, J.R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Philip, S.Y., Zhou, Z.H.: Top 10 algorithms in data mining. *Knowl. Inf. Syst.* **14**(1), 1–37 (2008)

7. Gruber, T.R.: A translation approach to portable ontology specifications. *Knowl. Acquisit.* **5**(2), 199–220 (1993)
8. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: From data mining to knowledge discovery in databases. *AI Mag.* **17**(3), 37 (1996)
9. Agrawal, R., Imieliński, T., Swami, A.: Mining association rules between sets of items in large databases. *ACM SIGMOD Rec.* **22**(2), 207–216 (1993)
10. Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., Verkamo, A.I.: Fast discovery of association rules. *Adv. Knowl. Discov. Data Mining* **12**(1), 307–328 (1996)
11. Frank, E., Hall, M., Holmes, G., Kirkby, R., Pfahringer, B., Witten, I.H., Trigg, L.: Weka. In: *Data Mining and Knowledge Discovery Handbook*, pp. 1305–1314. Springer US (2005)
12. “D2rq”. d2rq.org/. Accessed 24 June 2016
13. Lavrač, N., Vavpetič, A., Soldatova, L., Trajkovski, I., Novak, P.K.: Using ontologies in semantic data mining with SEGS and g-SEGS. In: Elomaa, T., Hollmén, J., Mannila, H. (eds.) *DS 2011. LNCS (LNAD)*, vol. 6926, pp. 165–178. Springer, Heidelberg (2011). doi: [10.1007/978-3-642-24477-3_15](https://doi.org/10.1007/978-3-642-24477-3_15)
14. Nebot, V., Berlanga, R.: Finding association rules in semantic web data. *Knowl. Based Syst.* **25**(1), 51–62 (2012)
15. Abedjan, Z., Naumann, F.: Improving RDF data through association rule mining. *Datenbank-Spektrum* **13**(2), 111–120 (2013)
16. Ahmedi, L., Sejdiu, B., Bytyçi, E., Ahmedi, F.: An integrated web portal for water quality monitoring through wireless sensor networks. *Int. J. Web Portals (IJWP)* **7**(1), 28–46 (2015)