Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 113 (2017) 360–367

Procedia Computer Science

www.elsevier.com/locate/procedia

The 7th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH 2017)

Enrichment of Association Rules through Exploitation of Ontology Properties – Healthcare Case Study

Eliot Bytyçi[a], Lule Ahmedi[a,*], Francesca A. Lisi[b]

[a]*University of Prishtina "Hasan Prishtina", Rr. "George Bush", p.n.,Prishtina 10000, Republic of Kosovo*
[b]*Dipartimento di Informatica, Università degli Studi di Bari "Aldo Moro", Via E. Orabona, 4, Bari 70126, Italy*

Abstract

Association rule mining as descriptive data mining category aims to find interesting patterns on data. The quality of the patterns is measured with two metrics: confidence and support. Especially in fields dealing with sensitive data, such as healthcare, the resulting patterns should be novel and interesting. To achieve that, not only the quality of the data itself should be superior, but also other additional attributes added, do support the results. That should be achieved by using Semantic Web technologies and thus enriching data used with semantic relations between properties. A hypothesis suggests that especially tackling property relations, chain property being part of the current version of the W3C Web Ontology Language (OWL), will yield better rules. To validate the hypothesis, experiments were performed on raw data, then on an older version of OWL, which does not support the chain properties and finally on the current version of language involving chain properties. Results obtained suggest that the latter produces novel rules with strong confidence and support, not encountered in former two experiments.

Peer-review under responsibility of the Conference Program Chairs.

Keywords: association rules, semantic web, property chains

1. Introduction

Data mining is seen as part of a bigger process which includes preprocessing, analyzing and summarizing mined knowledge for getting insight on the data related process[1]. Data mining tasks are divided into two categories:

* Corresponding author. Tel.: +381 38 554 896; fax: +381 38 42 525.
E-mail address: lule.ahmedi@uni-pr.edu

1877-0509 © 2017 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the Conference Program Chairs.
10.1016/j.procs.2017.08.345

predictive and descriptive. The predictive category, as the name suggest, aims to predict future feature significances. The descriptive category on the other hand, aims to find interesting patterns describing the data. The best representative of the descriptive category is association rule mining (ARM)[2]. One of the most used algorithms of the ARM is Apriori[3]. It is used for finding rules in different kinds of databases which contain transactions, such as the items bought together, medical records of a patient, etc. Different application domains dealing with data, that are eager to find relations between those data, use the Apriori and its derivatives.

One of those domains, involving usage of massive data, is healthcare. The most important portion of the healthcare data concerns patient medical history and treatment records. By analyzing the data, healthcare industry can unleash tremendous potential and usefulness[4]. Results of the analysis can be used for patient disease diagnoses, patient profiling or even history generation. All of those support medical staff in their everyday decision making. However, even though the technology has been tried and tested and proven right, still there are concerns regarding the data itself. One of the questions that arise is: are databases used unified for all different kind of sources?

Kolias et al.[5] believe that by using Semantic Web technologies the above mentioned concern could be overcome. Furthermore, the community would be able to retrieve hidden knowledge that remains unexploited in vast and diverse pools of medical data. Supportive to that is the fact that Semantic Web technologies provide tools to more accurately and effectively integrate and process data. One of the basic components of the Semantic Web are collections of information, called ontologies[6].

Additionally, ontologies could be used to infer new relations by usage of the advanced properties of the concepts described. On the basis of that fact, a hypothesis could be constructed that supports the claim that usage of advanced properties would led to new stronger and more confident rules. The hypothesis could be verified with experiments involving the application of ARM algorithms on several types of data. First, the ARM would be applied to raw patient data and secondly on ontology populated patient data but leaving aside advanced properties. In the end, experiments would be performed on ontology populated data enriched with advanced properties.

The rest of the paper is organized as follows: Section 2 describes the related work, Section 3 provides background on OWL2 and ARM, Section 4 presents the ontology created and used, Section 5 introduces the setup and the results of the experiments performed and Section 6 discusses concluding remarks and future research directions.

2. Related work

Abedjan and Naumann[7] suggest that ARM benefits from integration and usability of semantic data. One of the mining configurations mentioned is mining predicates. Mining predicates will result in gaining dependencies among elements of the schema or the properties of the subjects. In relation to our case, by introducing further object properties, we gain new relations between elements, which as will be shown, results into better rules.

Nebot and Berlanga[8] use schema level knowledge encoded in ontologies to derive appropriate transactions, which are then fed into traditional ARM algorithms. Experiments were performed on semantic data of biomedical application and showed usefulness and efficiency of the approach. Still, the approach required domain experts to define targets and contexts of the mining process, so that the correct transactions are generated.

Bytyçi et al.[9] use a different approach by enriching association rules with context ontologies. In order to evaluate their approach, they compare results obtained from deploying association rule mining in raw data and in context ontology derived data. Results inclined towards the context data association rule mining, present new rules not otherwise inferred following the other approach.

J´ozefowska et al.[10] presents a hybrid method of combining Semantic Web ontologies expressed in description logic as well rules in disjunctive Datalog. Ontology level association rules are extracted, presented as queries in Datalog. Used description logic SHIF[11] is the DL logic of OWL lite which then with DL-safe rules represents Semantic Web Rule Language (SWRL) rules but restricted to known individuals. Furthermore, the approach ignores mining of rules over predicates.

Lisi[12] performed research on mining the Semantic Web present in the fields of inductive logic programming (ILP) and generalization that make use of the description logic of a knowledge base. They concentrate on mining answer-sets of queries towards a knowledge base. Based on a general reference concept, additional logical relations are considered for refining the entries in an answer set.