

# Accepted Manuscript

Mining Semantic Association Rules from RDF Data

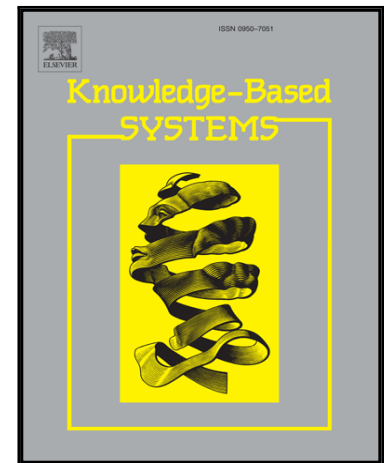
Molood Barati, Quan Bai, Qing Liu

PII: S0950-7051(17)30325-8  
DOI: [10.1016/j.knosys.2017.07.009](https://doi.org/10.1016/j.knosys.2017.07.009)  
Reference: KNOSYS 3973

To appear in: *Knowledge-Based Systems*

Received date: 14 July 2016  
Revised date: 30 June 2017  
Accepted date: 8 July 2017

Please cite this article as: Molood Barati, Quan Bai, Qing Liu, Mining Semantic Association Rules from RDF Data, *Knowledge-Based Systems* (2017), doi: [10.1016/j.knosys.2017.07.009](https://doi.org/10.1016/j.knosys.2017.07.009)



This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# Mining Semantic Association Rules from RDF Data

Molood Barati<sup>a,\*</sup>, Quan Bai<sup>a,\*</sup>, Qing Liu<sup>b</sup>

<sup>a</sup>*School of Engineering, Computer and Mathematical Sciences, Auckland University of Technology, Auckland 1010, New Zealand*

<sup>b</sup>*Software and Computational Systems, Data61, The Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia*

## Abstract

The Semantic Web opens up new opportunities for the data mining research. Semantic Web data is usually represented in the RDF triple format (subject, predicate, object). Large RDF-style Knowledge Bases contain hundreds of millions of RDF triples that represent knowledge in a machine-understandable format. Association rule mining is one of the most effective techniques for detecting frequent patterns. In the context of Semantic Web data mining, most existing methods rely on users intervention that is time-consuming and error-prone due to a large amount of data. Meanwhile, rule quality factors (e.g. support and confidence) usually consider knowledge at the instance-level. Namely, these factors disregard the knowledge embedded at the schema-level. In this paper, we demonstrate that ignoring knowledge encoded at the schema-level negatively impacts the interpretation of discovered rules. We introduce an approach called SWARM (Semantic Web Association Rule Mining) that automatically mines Semantic Association Rules from RDF data. The main achievement of SWARM is to reveal common behavioural patterns associated with knowledge at the instance-level and schema-level. We discuss how to utilize knowledge encoded at the schema-level to add more semantics to the rules. We compare the semantic of rules discovered by SWARM with one of the latest approaches in this field to show the importance of considering schema-level knowledge. Initial experiments performed on RDF-style Knowledge Bases demonstrate the effectiveness of the proposed approach.

**Keywords:** Semantic Web data, Association Rule Mining, Ontology, Knowledge Discovery

## 1. Introduction

The Internet can be considered as one of the most progressive inventions in the human communication area. Nowadays, the ever-growing amount of data on the Web has made it difficult to discover and analyze the information required by users. The Semantic Web (SW) is an effort to make knowledge on the Web both human-understandable and machine-readable [1]. SW data are usually structured in the triple format (subject, predicate, object) called Resource Description Framework (RDF). RDF relies on XML as an interchangeable data structure, creating an ontology for exchanging information on the Web. In recent years, by developing RDF/S, OWL, and SPARQL standardization, the number of large Knowledge Bases (KBs) available on the Linked Open Data cloud (LOD) has led to an explosive growth of the Web of data. RDF-style KBs such as DBpedia [2], YAGO [3], and Freebase<sup>1</sup> have a high contribution of distributing SW data. In July 2014, the number of triples in the LOD has exceeded 61 billion<sup>2</sup>.

Although these KBs suffer many issues such as incompleteness and incorrectness, they already contain millions

of triples that create new opportunities for data mining communities. Researchers have been recently working on developing tools which promise more potential for SW applications [4]. Although these tools facilitate the use of SW data, most of them provide traditional analytical techniques that are not sufficient for this kind of data. To overcome this issue, researchers have been using Data Mining (DM) techniques to analyze SW data. In this regard, association rule mining is one of the most common techniques for discovering frequent patterns.

There are several methods of mining associations from large RDF-style KBs. Inductive Logic Programming (ILP) is one of the well-known techniques for mining association rules from SW data [5]. ILP usually requires counterexamples. AMIE [6, 7] is a multi-threaded approach where the Knowledge Base (KB) is kept and indexed in the memory. High memory usage is one of the drawbacks of AMIE. This method concentrates on mining Horn rules among predicates such as  $motherOf(m,c) \wedge marriedTo(m,f) \Rightarrow fatherOf(f,c)$ . AMIE's quality factors (support and confidence) only consider knowledge at the instance-level and remove the *rdf:type* relations from the KB (i.e., schema-level or ontology). In order to reduce the number of generated Horn rules, the approach mines closed rules, i.e., each of the variables in the predicates should have appeared at least twice. For example, in the above rule, the variable *m* has been shared between *motherOf* and *marriedTo* twice.

\*Corresponding author: Tel.: +64 9 921 9999; Fax: +64 9 921 9812.

Email addresses: mbarati@aut.ac.nz (Molood Barati),

qbai@aut.ac.nz (Quan Bai), Q.Liu@data61.csiro.au (Qing Liu)

<sup>1</sup><http://freebase.com>

<sup>2</sup><http://stats.lod2.eu>

Download English Version:

<https://daneshyari.com/en/article/4946109>

Download Persian Version:

<https://daneshyari.com/article/4946109>

[Daneshyari.com](https://daneshyari.com)