Using Neural Networks to Aggregate Linked Data Rules

Ilaria Tiddi, Mathieu d'Aquin, and Enrico Motta

Knowledge Media Institute
The Open University, United Kingdom
{ilaria.tiddi,mathieu.daquin,enrico.motta}@open.ac.uk

Abstract. Two typical problems are encountered after obtaining a set of rules from a data mining process: (i) their number can be extremely large and (ii) not all of them are interesting to be considered. Both manual and automatic strategies trying to overcome those problems have to deal with technical issues such as time costs and computational complexity. This work is an attempt to address the quantity and quality issues through using a Neural Network model for predicting the quality of Linked Data rules. Our motivation comes from our previous work, in which we obtained large sets of atomic rules through an inductive logic inspired process traversing Linked Data. Assuming a limited amount of resources, and therefore the impossibility of trying every possible combination to obtain a better rule representing a subset of items, the major issue becomes detecting the combinations that will produce the best rule in the shortest time. Therefore, we propose to use a Neural Network to learn directly from the rules how to recognise a promising aggregation. Our experiments show that including a Neural Network-based prediction model in a rule aggregation process significantly reduces the amount of resources (time and space) required to produce high-quality rules.

Keywords: Linked Data, Rules Aggregation, Neural Networks.

1 Introduction

When running a Knowledge Discovery (KD) process, the main ambition is to provide humans with useful, explicit information about the collected data. Such information, called *nugget* or *pattern*, is a statement describing an interesting relationship among a subset of the data [8]. Patterns are to be considered as any result of a data mining process, such as a series of clusters, a classification model, rules, etc. Two typical problems are encountered when producing patterns from a data mining process: (i) their number can be extremely large (data quantity problem) and (ii) not all of them are interesting to be considered (rule quality problem).

The typical strategy to overcome those problems is providing the patterns to human experts, whose role consists in analysing the results, discovering the interesting ones while explaining, pruning or refining the unclear ones. To cope

K. Janowicz et al. (Eds.): EKAW 2014, LNAI 8876, pp. 547–562, 2014.

with such a strenuous and time consuming process, many strategies to automatically assist the experts in the post-mining phase have been proposed in the past years, ranging from using interestingness measures for rules [10] to including the ontological knowledge [1,14,16,21]. However, technical problems are often to be encountered, such as time costs and scalability of the proposed processes.

In particular, those issues were encountered in our previous work, presented in [23], in which we described an Inductive Logic Programming-inspired framework to give Linked Data based explanations (rules applying to specific patterns – see Section 3), for clusters automatically obtained out a data mining process. While the obtained explanations resulted to be interesting and representative for each cluster, our major issue consisted in having to deal with a large amount of atomic rules (rules made of a single condition in the body) that need to be aggregated in a post-processing step in order to obtain a better representation of our clusters. When having to deal with such large amounts of rules, trying each possible aggregation requires too many resources. Most of the aggregations between rules might not be interesting, i.e. they would not produce any rule better representing the cluster, while the time and computational costs required by the process increase exponentially with the amount of rules to aggregate. To reduce such a process, what is required is a way to predict which rules are worth aggregating. However, in order to do so, it is first important to detect what knowledge about the rules is required to make a prediction.

In this work, we present a Neural Network-based approach to predict whether two rules, if combined, will lead to a better rule. We use statistical information about the rules (precision, recall and F-measure), as indicators for a prediction. Finally, we compare our process to other strategies for rules aggregation. The rest of the paper is articulated as follows. Section 2 presents the related work. Section 3 summarises the framework presented in [23] to produce Linked Data-based explanations, the motivation behind it and the unsolved questions that it left. Section 4 presents our problems, how we end up with proposing a Neural Network-based approach and the process we designed. Section 5 presents the experiments we ran, comparing our approach to some intuitive strategies for rules aggregation. The last section concerns future work and conclusion.

2 Related Work

No previous work seems to have been published in which Neural Networks are applied to Linked Data structures. Below we present some fields connecting to our research.

Association Rules Mining and Artificial Neural Networks. The problem of dealing with many rules and how to detect the interesting ones is a typical issue of Association Rule Mining (ARM). Many researchers have proposed methods for automatically detecting rules interestingness in the past. An extensive survey can be found in [10].