Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 96 (2016) 345 – 354

Procedia Computer Science

20th International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2016, 5-7 September 2016, York, United Kingdom

Ontology Knowledge mining based Association Rules Ranking

Rihab Idoudi[a,b*] , Karim Saheb Ettabaa [b], Basel Solaiman [b],Kamel Hamrouni [a]

[a] Université Tunis ElManar, Ecole Nationale d'Ingénieurs de Tunis, Tunis, 1200, Tunisia
[b] ITI Laboratoire, Telecom Bretagne, Brest, 29238, France

Abstract:

Medical association rules induction is used to discover useful correlations between pertinent concepts from large medical databases. Nevertheless, ARs algorithms produce huge amount of delivered rules and do not guarantee the usefulness and interestingness of the generated knowledge. To overcome this drawback, we propose an ontology based interestingness measure for ARs ranking. According to domain expert, the goal of the use of ARs is to discover implicit relationships between items of different categories such as 'clinical features and disorders','clinical features and radiological observations', etc. That's to say, the itemsets which are composed of "similar" items are uninteresting. Therefore, the dissimilarity between the rule's items can be used to judge the interestingness of association rules; the more different are the items, the more interesting the rule is. In this paper, we design a distinct approach for ranking semantically interesting association rules involving the use of an ontology knowledge mining approach. The basic idea is to organize the ontology's concepts into a hierarchical structure of conceptual clusters of targeted subjects, where each cluster encapsulates "similar" concepts suggesting a specific category of the domain knowledge. The interestingness of association rules is, then, defined as the dissimilarity between corresponding clusters. That's to say, the further are the clusters of the items in the AR, the more interesting the rule is. We apply the method in our domain of interest – mammographic domain- using an existing mammographic ontology called Mammo*, with the goal of deriving interesting rules from past experiences, to discover implicit relationships between concepts modeling the domain.

Peer-review under responsibility of KES International

Keywords: Association Rule, ontology knowledge mining, conceptual clusters, interestingness measures, Ranking

* http://sourceforge.net/p/gimimammography/code/HEAD/tree/trunk/owl

Peer-review under responsibility of KES International
doi:10.1016/j.procs.2016.08.147

1. Introduction

Association rule mining is, actually, one of the most important tasks in knowledge extraction in databases [1]. This technique aims to discover implicit correlations between items in databases that can be of great interest to domain experts. A typical and widely used application of AR mining is the medical domain. In fact, the exponential increase of the volume as well as the complexity of the radiological data raises crucial needs, for data management and knowledge discovery. However, the large number of extracted association rules makes the task of processing and interpreting them, very complicated for domain experts. Addressing this issue, researchers have proposed to rank generated rules according to their potential interest and enables highly ranked rules to be straightaway presented to decision makers. Methods for interestingness rules examination can be divided into objective and subjective methods. In subjective analysis, the ARs evaluation is based on prior domain knowledge that can be modeled within different supports (such as Rule schemas, ontologies, etc.). While, the objective methods are based on the use of statistical information in the database. Although the latter method can bring useful information regarding the dataset structure such as the support and the confidence measures, the rule interestingness strongly depends on the domain knowledge. Indeed, the more the knowledge is represented in an expressive and formal way, the more the rule evaluation is efficient. To this end, some approaches have proposed to involve ontologies in the post-processing task, to model domain knowledge since they provide a semantic support for vocabularies structuring. Recently, in medical domains, ontologies have become the cornerstone in knowledge acquisition and formalizing [2]. For example, in the mammographic domain, the mammographic ontology might be used to describe: the radiological observations associated with feature descriptors, mammogram Bi-Rads classification, clinical observation, etc. Surprisingly, while intense researches exist on applying association rule to mine ontologies, few approaches have, so far, exploited the benefits brought by these ontologies to compute rule interestingness measures [2]. Such measure represents a function that takes ontology concepts as input and returns a numerical value reflecting the dissimilarity between these concepts. Generally, this conceptual dissimilarity (distance) is based on the use of path based measure. That's to say, the furthest the concepts are in the ontological hierarchy, the more interesting the rule is [3] [2].

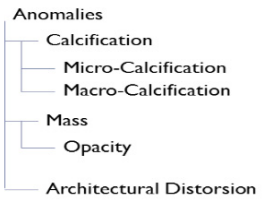

Figure 1: Extract of mammographic ontology

For example, the distance between the ontological concepts 'Calcification' and 'Micro-calcification' would be interpreted differently from the distance between 'Calcification' and 'Opacity' (see Figure 1), because of their different localization regarding the root. Moreover, the computed distance is strongly limited to the subjective construction of the ontology. Basing on multiple interviews, domain expert have revealed that these concepts like any two other entities regarding the anomalies category (see Figure1) are to a certain extent semantically similar since they are a part of a single subject. In order to meet the constraints imposed by the domain experts, we propose in this paper to increase the abstraction level of the conceptual knowledge encoded in the ontology, through the ontology knowledge mining process. The proposed idea consists on mining a novel structure of conceptual clusters organized hierarchically [24]. Each cluster groups similar objects encoding a specific topic regarding the ontology knowledge. That's to say, we aim to investigate the mined hierarchical conceptual clusters to determine so-called 'semantic interestingness measures' by computing the dissimilarity between the items of a given rule. Replacing exact similarity measure between concepts with semantic similarity between their correspondent clusters enables novel ways of interpreting AR, and hence may lead to the right identification of the interestingness of the rules. Therefore, the more clusters are far away, the more dissimilar are their respective concepts. The use of efficient data mining methods and appropriate interestingness metrics enables the identification of high quality relationships [3].