Association rules-based Ontology Enrichment

RihabIdoudi[1, 2], Karim Saheb Ettabaa[2], Basel Solaiman[2], Kamel Hamrouni[1], Najla Mnif [3]
[1]Université Tunis ElManar
Ecole Nationale d'Ingénieurs de Tunis
Tunis, 1200, Tunisia
[2]ITI Laboratoire, Telecom Bretagne
Brest, 29238, France
[3]Hospital Charles Nicoles

ABSTRACT: *Among the most powerful tools for knowledge representation, we cite the ontology which allows knowledge structuring and sharing. In order to achieve efficient domain knowledge bases content, the latter has to establish well linked and knowledge between its components. In parallel, data mining techniques are used to discover hidden structures within large databases. In particular, association rules are used to discover co-occurrence relationships from past experiences. In this context, we propose, to develop a method to enrich existing ontologies with the identification of novel semantic relations between concepts in order to have a better coverage of the domain knowledge. The enrichment process is realized through discovered association rules. Nevertheless, this technique generates a large number of rules, where some of them, may be evident or already declared in the knowledge base. To this end, the generated association rules are categorized into three main classes: known knowledge, novel knowledge and unexpected rules. We demonstrate the applicability of this method using an existing mammographic ontology and patient's records.*

Keywords: Ontology, Association rules, Association rules Categorization, Known knowledge, Novel knowledge and Unexpected rules

Received: 13 December 2015, Revised 12 January 2016, Accepted 19 January 2016

1. Introduction

Ontology is a knowledge representation tool which allows shared understanding of a given domain in order to improve communication among humans and machines. It defines the semantics of a field through representing a corpse of formally represented knowledge holding the concepts, and their relationships[1]. Because of that, ontology has become a crucial tool that aims to agree a comprehension of a domain.

Nevertheless, for several domains, knowledge is evolving in spectacular way in terms of the number of elements or relations among them. In this respect, respective ontologies are continuously subject to regular updates and developments. Performing these updates manually remains a costly and time consuming task since it requires mobilizing one or more domain experts to identify and then, classify the new vocabulary in the ontology.

In parallel, Association rule (AR) mining is one of the most important data mining techniques commonly used for discovering hidden associations between data elements in a diverse range of applications [2]. A typical and widely used application of AR mining is medical domain. In fact, there has been an exponential increase in the volume and complexity of available radiological data, which needs to be addressed with the methods of data management and knowledge discovery so to be further employed.

In this paper, we propose to exploit the potential usefulness of past experiences, in order to supply in an automatic way, existing knowledge bases with new semantic associations. Such enrichment method, allows drawing a better comprehension of the domain. However, the generated ARs may encapsulate already known knowledge, where we can state, it is crucial to compare such a derived knowledge to the existing one. The novelty of our approach is that it addresses the coupling of ontology and ARs; so it allows categorizing the generated ARs into three main classes. **(i) Known rule:** they will not be further employed, **(ii) Novel rules:** this category is of great interest to the domain expert since it allows adding new relations between concepts in the ontology, **(iii) Unexpected rules:** those rules contradict the modeled knowledge in the ontology. Both latter categories are straightaway presented to the domain expert to be validated.

Knowing process is the major challenge of knowledge management which aims to conceive useful and meaningful domain knowledge. We use this concept for deriving novel knowledge with the use of data mining techniques. We focus, in this paper, on the mammographic domain, where our objective is to discover implicit relationships between mammography's concepts such as 'clinical features and disorders', 'clinical features and radiological observations', etc. The overall approach starts, first by extracting ARs from the database. The latter comprises transactions of patients' medical records (previously diagnosed). In the next step, the process compares the derived knowledge to the existing knowledge in the ontology using a set of rule mining operators proposed in the literature. The mammographic ontology used in this paper is the 'Mammo' ontology which is open source and available in the net. Once, the interesting knowledge is validated by the domain expert, we process to enrich the ontology of interest with different relationships.

The paper is organized as follows. First, we advance the state of the art by reconnoitering the current approaches for ontology learning. Second, we introduce the general definitions of basic notions used in this manuscript. Third, we introduce our approach for the ontology enrichment basing on the use of AR.

2. Related Work

According to [3], few works have addressed, so far, the coupling of ontology and association rules issue. While, most of them have been interested to ontology mining which is a process for learning an ontology (including classes, classes hierarchy and property)[4] [5] [6] [7] [8] [9] [10], few among them have reflected the ways to enrich existing ontology by supporting the creation of semantic relations between ontology concepts.

In[11], authors have defined the ontology enrichment process as two main steps: a learning step (which consists on new concepts and relations) and a placement step (which consists on finding the appropriate place in the original knowledge base). They focused on new ways to discover new terms as well as relations with initial concepts in the ontology of interest.

In[12], authors mined associations between medical concepts from medical textual documents related to breast cancer treatment by the means of ARs. The work focuses on finding specific semantic relations for the associated concept pairs. Similarly, authors in [13] focus on mining the ontologies from text documents (e.g., web content) or other web data (web usage, web structure and web user profiles). AR mining has been adopted to discover the relationships between different concepts in [14]. The proposed approach in [15] consists on extracting relevant terms from document, in order to enrich the ontological model with relations between concepts. They have used methods for relevant term extraction from documents, the algorithm FP-Growth to collect frequent itemsets and association rule learning to learn useful relations from frequent item sets.

In[16], authors apply ARs on manually detected concepts with the goal of concept enrichment, i.e. discovering new concepts. The use of ARs is justified by the fact that they lead to the derivation of useful associations and correlations within data.

In[17], authors proposed to exploit the evidence coming from the data for discovering hidden knowledge patterns to be used for extending ontologies with formal rules and suggesting new knowledge axioms. The idea is to transform generated patterns into rules or axioms where atoms may designate concepts or roles. Authors employed as well the notion of operators so that connected and non-redundant rules are obtained.

Most of the proposed approaches have been interested in learning ontologies from scratch which enable for discovering relevant concepts and relations from corpus of data (text documents, web content) with the use of data mining techniques. In this paper, we are exploiting novel knowledge coming from medical records in order to evolve the content of an existing mammographic ontology. That's to say, considering an existing ontology, we compare the existing associations, which have been already defined in the knowledge base, and the novel knowledge coming from discovered associations rules with the means of Liu operators. The output of the comparison step is a set of categorized rules: known rules, unexpected rules and novel rules. Unexpected as well as novel rules are of great interest to domain experts. Once these rules are validated by the expert, we proceed to enrich the knowledge base with association between the exiting concepts.

Our approach is inspired from the work proposed in[18], where the comparison process is carried rule schemas (in order to model the user expectations) and ARs. The objective of this work is to filter ARs that are directly mapped to defined rules schemas using Liu operators [19].

3. Basic Notions

3.1 Association Rule Mining

In data mining process, ARs are used for discovering important relations between items in a database D, where $D = \{t, t_2, t_m\}$ is a set of transactions over $I = \{i, i_2, i_n\}$ which is a set of items. A non-empty subset of I, $X = \{i, i_2, i_k\}$ is called an itemset. Each transaction t_i in D is defined as an itemset i, $i_2,..., i_k$ of length k.

An *AR* is an implication between two itemsets X and Y, in the form of $X \rightarrow Y$ where $X \cap Y = \emptyset$, that's to say they do not have any items in common. X is called the antecedent and Y is the conclusion, or the consequent, of the association rule.

Each *AR R*: $X \rightarrow Y$ may be characterized by two measures Support and confidence. They are used for selecting ARs according to their potential interest to the user:

• The Support (sup) is defined as the occurrences of a specific event containing $X \cup Y$, $sup(X \rightarrow Y) = s$; s% of the transactions in D contains X and Y;

• The Confidence (conf) is defined as: $conf(X \rightarrow Y) = sup(X \rightarrow Y) / sup(X) = sup(X \cup Y) / sup(X) = c$; that's to say, c % of the number of transactions containing X contains as well as Y.

The algorithm Apriori[19] is the most widely used algorithm to discover ARs in databases. It gets as input; the database to be mined and two thresholds minConf and minSup which represent respectively the minimum values of confidence and support that an AR must hold. The algorithm consists of two main steps: First all the item sets where the support is greater than minSup are generated; second, rules with support and confidence greater than minConf and minSup are extracted from the item sets (generated in the first step).

2.2 Ontology Definition

An ontology is defined as a formal specification of a shared conceptualization. Ontologies have been used to capture and formalize knowledge by modeling concepts and relations associated to the domain of interest. Concepts designate the pertinent entities, for example, *mammogram* is a concept within the mammographic domain, whereas relations designate the interactions between revealed concepts, for example, mammogram *classified_as* Bi-Rads3(Breast Imaging Reporting and Data System), that's to say, the concepts *mammogram* and *Bi-Rads3* are related via the relationship *classified_a*. Relations in the ontology are categorized into:

• **Subsumption:** Used to define the taxonomy which refers to the hierarchical concept tree, for example Opacity is a type of anomalies(see Figure 1)
• **Associative relations:** Relate the different concepts of the hierarchy (e.g. *classified_as*).
• **Ontologies may define as well:** Individuals which represent instances of generic declared concepts and axioms used to constrain interpretations for classes, relations or instances.

Ontologies have been widely used in medical domains to capture knowledge and formalize medical lexicons. In the mammographic domain, several ontologies have been proposed such as BCGO[1], Mammo[2], Radlex[3]. Each ontology has been developed by different communities and for specific intended task.

In this paper, we propose to use the Mammo ontology since it is considered as well structured and rich of vocabularies (concepts and relationships) that are relevant to the mammographic domain such as anomalies description, diagnosis and mammogram classifications. The ontology is developed with the description logic formalism. The latter is characterized by frame-type sortal hierarchies and logical assertions, where sortal hierarchy is corresponding to classes and subclasses in the object-oriented model.

4. The Proposed Approach

The proposed approach is based on two main components, both of which are associated to the mammographic domain: (i) an existing mammographic ontology called Mammo formalizes the mammographic domain knowledge and (ii) The ARs discovered from the database. The latter includes different patient records obtained from radiological units. The generation of ARs is realized with the use of A-priori algorithm where the support and the confidence measures are adjusted by the user. As shown in Figure 1, the main steps of the approach are as following:

- Representation of the ontological associations in a rule like formalism.

- Extraction of the ARs using the Apriori algorithm with respect to the support and confidence thresholds

- Mapping AR's items to the ontological concepts.

- Categorization of the ARs into three main categories (Known rules, unexpected rules and novel rules) with the application of operators.

- Validation of the novel rules by the domain expert.

- Enriching ontology with novel validated associations between the existing concepts.

4.1 Ontology Knowledge Representation

In first place, we propose to transform the relations declared in the knowledge base in a set of obvious rules within a formalism similar to the AR representation A→B; where A and B are two ontology's concepts. This allows the comparison of both knowledge sources. The obvious rules set describe already known knowledge modeled in the ontology. As a result, those ontology rules act as a rule grouping, defining rule families. Each rule encloses the associated concepts behaving as items and '→' designates the implication between them. The implication may describe the subsumption relationship (used to build the hierarchy) or the associative ones (used to build affirmations about relationships between ontology concepts).

The generated obvious rules are two-items, that's to say they express correlations between two concepts. For example, the ontological relation <*Micro-calcification is-a calcification*> is converted to an obvious rule: *Micro-calcification→calcification.*

4.2 ARs Extraction

In the other hand, the AR mining algorithm is performed over the dataset with respect to predefined thresholds of support and confidence with the use of the Apriori algorithm. We are, particularly, interested to two-item rules of the generated ARs.

4.3 Mapping Ontology's Concepts and AR's Items

In order to perform comparison process, it is fundamental to connect the ontology's concepts to the database's items, each

[1]https://bioportal.bioontology.org/ontologies/BCGO

[2]http://sourceforge.net/p/gimimammography/code/HEAD/tree/trunk/owl/mammo.owl

[3]http://bioportal.bioontology.org/ontologies/RADLEX/?p=classes&conceptid=http%3A%2F%2Fwww.owl-ontologies.com%2FOntology1447432460.owl%23

one of them being connected to one or more items. To this end, we distinguish two types of ontology concepts: basic concepts and general concept. Basic concepts represent the set of concepts that do not subsume other concepts in the ontological hierarchy, whereas the general concepts designate those that subsume other concepts in the ontology.

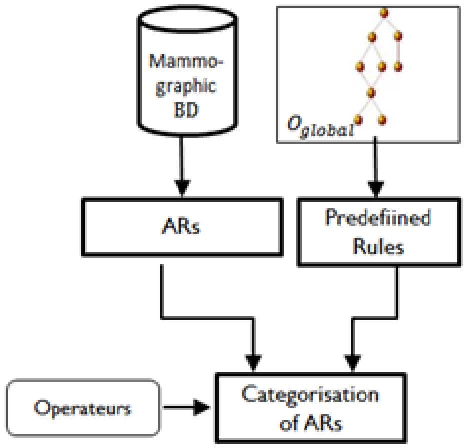

Figure 1. Approach Description

A basic concept is connected to the database's items through direct connection whereas a general concept is connected through direct and indirect concept. The direct connection ensures the mapping between equivalent entities, for example: $f_{direct}(mass\_oval) = mass\_oval$.

A general concept is connected to the database's items through direct connection and indirect connection. The latter ensures mapping ontology concepts to nearest items of the database through a generalization relation. For example, let us consider the concept *mass* which is a general concept. The latter is connected through an indirect connection to $mass_{oval}$. That's to say, if we consider the ontology rule *mass* $\rightarrow$ *benign_diagnosis* and the association rule: *mass_oval*$\rightarrow$ *benign_diagnosi*, through the mapping process, both rules are equivalent.

4.4 ARs Categorization

The ARs processing basing on the ontology knowledge is realized with the help of rule mining operators, allowing the categorization of the discovered rules into three main classes. To this end, we propose to reuse the operators proposed by Liu et al.: conforming and unexpectedness. The latter have been used in [18], in order to prune the AR with respect to predefined rule schemas that describe the user expectations.

The *conforming operator* Op_C: Applied over an ontology rule, extract the AR that is equivalent. That's to say, an AR is selected if its items in the condition and the conclusion match the concepts of the ontology rule. As a result, this operator allows discovering obvious rules that are already part of the ontology. The output of the application of this operator is a set of *known* rules.

The *unexpectedness operators* $Op_{U(C)}$, $Op_{U(A)}$: This kind of operator selects the AR where the item in the antecedent or the conclusion is, respectively, different from the concept in the antecedent or the conclusion of the ontology rule. The output of the application of this operator is a set of rules which are *unexpected* either regarding the antecedentor regarding the consequent.

The rest of the rules which were bot selected with none of the two operators constitute the set of the *novel* rules which are different with regard to the prior knowledge (See Figure 2).

Both categories of unexpected and novel rules are of great interest regarding the known rules since; we seek through this work to discover new knowledge to enrich the domain knowledge base. Nevertheless, these rules need to be evaluated in term of pertinence. To this end, these rules are straightaway presented to the domain expert in order to be analyzed and validated.

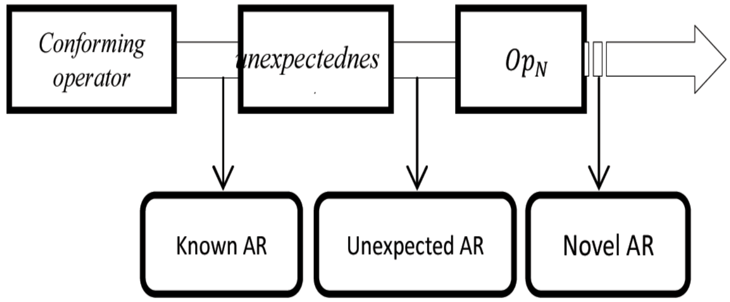

Figure 2. Rules Categorization

4.5 Validation of the Novel Rules

The validation of the unexpected as well as novel rules is realized manually by the domain expert. Among those rules, only some of them are of real interest and will be submitted to the enrichment process of the knowledge base. In this step, the expert may, permute the antecedent and the conclusion of the rule.

4.6 Ontology Enrichment

Our enrichment process aims to add relations between the concepts in the original ontology to establish well linked knowledge base (as we have mentioned above, we are only interested to the relational enrichment of the ontology).

Considering the items of the antecedent and the premise of the valid AR, we process either to create novel associative relation or use an existing one (This allows to avoid adding redundant links). Labeling the relationship (modeled within the implication) depends to the semantic meaning between the antecedent and the conclusion of the rule.

5. Results

5.1 Mammographic Ontology Description

We have selected the mammographic ontology 'Mammo' so to be used as a semantic support for ARs processing. This is due to the fact that it is rich of vocabularies and it is well structured according to[20]. This ontology has 692 concepts and 73 properties.

5.2 Dataset

Automated radiological systems are accumulating, daily, large quantities of information about patients and their medical conditions. Therefore, we propose to explore in this work, the medical data-repository which includes the patients' medical records previously diagnosed. The overall database includes 1000 patient records. Those data have been collected from the hospital Charles Nicolle in Tunisia. Each record encloses a textual description about, patient's information (such as the age, menopause, etc.), clinical observation (breast nature, skin change, etc.), radiological observation (tumor description) and the evaluation and/or mammogram classification (according to Bi-Rads classification).

The main objective is to find out relations between those classes such as correlations between clinical features and disorders, clinical features and radiological observations, clinical features and mammogram classification, etc.

In data processing, we have eliminated information that is not relevant to our context, and converted the past experiences into transactions with readable format (arff file) including 30 attributes values (Table 1 shows an extract of the used items).

5.3 Ontology Prior Knowledge Representation

The predicates introduced in the ontology are converted into a set of ontology rules. Table 2 shows an extract of the set of the obvious knowledge, where each ontology rule is associated with its semantic interpretation provided by an expert.

5.4 The ARs Extraction

The Apriori Algorithm is used here for AR Mining to extract rules that satisfy the predefined minimum support and confi dence. We have chosen the Weka3.6.1 [4]software which is an open source data mining software.

In order to observe the influence of support and confidence thresholds on the generated rules, we have experimented different parameter thresholds settings. Table 3 shows the numbers of generated rules with the variation of MinSup and Minconf parameters.

Attributes	
Breast Neoplasm	biopsy
Breast Carcinoma	Benign_diagnosis
Bi-Rads 1	Malin diagnosis
Bi-Rads 3	Follow-up
Bi-Rads 4	Nipple retraction
Mass_oval	Normal_breast
Mass-round	fibrocystic_breast
Irregular_mass	Breast pain
Parallel_mass	Age[60-69]
Low_density	Mastalgia
Not_defined_margin	Normal_breast
Circumscribed_margin	fibrocystic_breast
Speculated_margin	Breast pain
Speculated lesion	Age[60-69]
Skin retraction	Mastalgia

Table 1. An extract of the dataset attributes values

Predicates	Ontology Rule	Semantic interpretation by expert
Associated_with (mass_oval, benign_diagnosis)	mass_oval→ benign_diagnosis	A well oval massis highly predictive of benign l
Associated_with(irregular_mass, malign_diagnosis)	irregular_mass→ malign_diagnosis	Irregular mass shape is highly predictive malignant mass.
Associated_with(parallel_mass, benign_diagnosis)	parallel_mass→benign_diagnosis	Parallel orientation with respect to the skin surface is highly predictive benign mass.
Associated_with(fibrocystic_breast, normal_breast)	fibrocystic_breast → normal_breast	Fibrocystic breast lumps are completely benign, and are not associated with any risk for the future development of breast cancer.
determine(Breast_pain, mastalgia)	Breast_pain→mastalgia	Breast pain is be classified as cyclic mastalgia or non-cyclic Mastalgia
Detect(palpation, tumor_form)	palpation→tumor_form	Abdominal palpation can detect the tumor form
Require(Bi-rads3, follow_up)	Bi-rads3 →follow_up	BI-RADS category 3 lesions are recommended for 6-month follow-up

Table 2. Predicates generated from the ontology

Itération	1	2	3	4	5	6	7	8	9	10
MinSup	0.7	0.3	0.1	0.7	0.3	0.1	0.1	0.09	0.07	0. 05
MinConf	1	1	1	0.9	0.9	0.9	0.8	0.7	0.7	0.7
RA	0	0	74	2	23	536	1500	2167	3395	6552

Table 3. Numbers of rules generated for various minimum supports

It can be seen that high *minConf* and *minSup* values lead to fewer but more robust rules, i.e. rules with a high conditional probability, antecedent and consequent being almost always correlated, while decreasing the minimum support as well as confidence values, can conspicuously increase the number of rules.

---

[1] http://sourceforge.net/projects/weka/files/weka-3-7-windows-x64/

As a matter of fact, if the parameters of interest, i.e., the threshold values, are fixed too high, then too few rules are generated with omitting useful information. Otherwise, if the parameters are fixed too low, then, the algorithms can generate an extremely large amount of rules with unsuitable or uninteresting knowledge.

To find a good compromise between number of rules and robustness, we have emprically chosen *minSup* = 0.1 and *minConf* = 0.8 leading to the extraction of 1500 rules, where 734 rules are two-items.

5.5 ARs Processing

Extracted ARs are grouped in a rules file, and then two iterations were performed above those rules with the application of operators separately. In the first iteration, the conforming operator was applied Op_C, in second step $Op_{U(C)}$ and $Op_{U(A)}$ are performed.

First, the confirming operator filters already 500 rules known rules namely 68% of ARs. Although this category of ARs is not of great interest for experts and does not bring important information, it indicates the high coverage rate of the ontology knowledge regarding the mammography domain. The original rule set has been subsequently pruned by discarding the set of obvious knowledge. Next, the unexpectedness operator is applied to discover only 10 unexpected rules namely, 1% of ARs. This low percentage indicates as well the high reliability rate of the ontology. Finally, 224 rules are selected as novel knowledge, namely 31%. A close examination on ontology rules, which were not discovered during ARs generation, revealed that the used dataset is not full representative and need to be enlarged with more cases.

Original AR	Known AR	Unexpected AR	Novel AR
734	68%	1%	31%

Table 4. AR categorization

Table 5 shows an extract of novel rules set with correspondent confidence and support.

Therefore, the rules quality has to be analyzed and verified by the expert; in fact of matter, some rules may be either uninteresting and do not bring important information for experts, or even they may be inconsistent regarding domain knowledge. For example:

$$C13 \rightarrow C7$$
Support = 88%; Confidence = 80%

This can be translated to ‘a Lesion with a speculated margin has a low density’. This rule has no signification for the expert and cannot be useful for the knowledge domain. However, some surprising rules can provide additional meaning, improve the interoperability and comprehensibility for the expert and therefore need to be further studied, for example, this rule:

$$C4 \rightarrow B4$$
Support = 71%; Confidence = 83%

This can be translated to ‘round form is associated with the category Bi-rads4’. This rule is very interesting for domain expert because it is usually common that the form or the margin of the lesion is an indicator of malignancy, but it is not related to Bi-Rads classification.

N	ARs	confidence	Support
1	Mass_oval→ Bi-Rads 4	83%	48%
2	E10→Parallel_mass	83%	75%
3	Circumscribed_margin→ Bi-Rads 4	80%	22%
4	Mass-round→ Bi-Rads 4	83%	69%
5	Breast pain→E10	88%	49%
6	Age[60-69]→Parallel_mass	88%	88%
7	Not_defined_margin→Parallel_mass	86%	65%

Table 5. Novel rules

Another experiment was performed to visualize the variation on the novel rules number with respect to minSup. Figure 3 plots the results of the experiment. It is well notable that while increasing the minSup, the number of rules decreases.

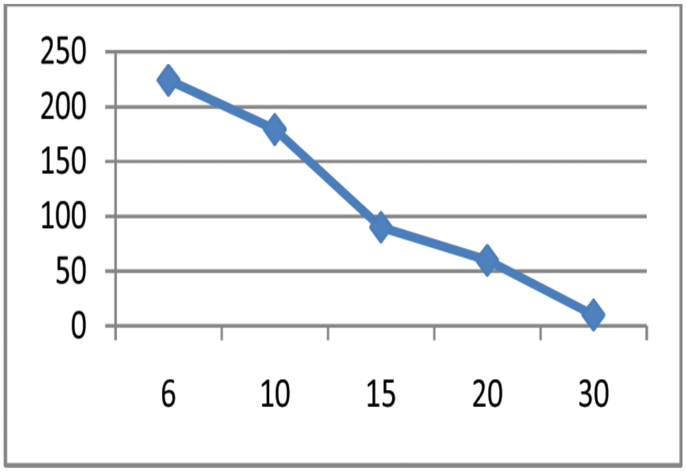

Figure 3. Comparison of novel rules numbers according to different minSupport thresholds

5.6 Ontology Enrichment

This step ensures ontology enrichment with the novel knowledge extracted from the ARs after being examined and validated by experts. The latter may recall existing object properties (Associated_with(), Require()…) in the ontology or may enrich the ontology with new and appropriate ones to create the novel predicates. Among the new relations added to the ontology, we have:<<Topologically_connected_to>>: designates the way the constituents are interrelated; <<Produce>>: A causative agent or process leading to the appearance of another. An example of a new predicate: <Skin retraction, Associated_with, Mastalgia>. As a result, the ontology has been enriched with 50 relationships between the existing concpets.

6. Conclusion

In this paper, the developed work lies within the scope of domain ontologies relational enrichment from anterior experiences using data mining techniques. Although several approaches have been interested to ontology learning from scratch, few works have focused on evolving the ontology content by adding different knowledge. The proposed approach takes advantage from knowledge extracted from existing database in order to be further employed and shared. To this end, we have first confronted both knowledge sources by converting the predicates in the ontology in a rule like formalism. The different novel knowledge have been examined and analyzed by the domain expert. The rules as they express correlations between the concepts, relations have been added in the ontology of interest. In this paper, we have dealt with the mammographic domain, where the used database includes patient records collected from the hospitals, and the used mammographic ontology is the mammo ontology.

References

[1] Ahmed, E. B., et., Gargouri, F. (2015). Enhanced Association Rules over Ontology Resources, *International Journal Web Applications*, 7 (11) 10-23.

[2] Mohd, I. N., Hadzic, S. F., et Dillon, T. (2011). Interestingness measures for association rules based on statistical validity, *Knowledge-Based Systems*, 124, 386–392.

[3] Ahmed, E. B., et., Gargouri, F. (2015). Enhanced Association Rules over Ontology Resources, *International Journal of Web Applications*, 7 (11) 10-22.

[4] Wong, W., Liu, W., et., Bennamoun, M. (2012). Ontology learning from text: A look back and into the future, *ACM Computing Surveys* (CSUR) 4. 144, 20.

[5] Dou, D., Frishkoff, G., Rong, J., Frank, R., Malony, A., et., Tucker, D. (2007). Development of Neuro Electro Magnetic ontologies (NEMO): a framework for mining brainwave ontologies, *In:* Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data mining, 270-279.

[6] Bellandi, A., Furletti, B., Grossi, V., et., A Rome. (2007). Ontology-driven association rules extraction: a case study,» Workshop Context and Ontologies: *Representation and Reasoning*, 1-10.

[7] Kuo, Y., Lonie, A., Sonenberg, L., et., Paizis, K. (2007). Domain ontology driven data mining – a medical case study, *International Conference on Knowledge Discovery and Data Mining*, 11–17.

[8] Phillips, J., et., Buchanan, B. G. (2001). Ontology guided knowledge discovery in databases, *First International Conference on Knowledge Capture*, 123–130.

[9] Sharma, S. et., Osei-Bryson, K. (2008). Organization-ontology based framework for implementing the business understanding phase of data mining projects, 41st *Hawaii International Conference on Systems Sciences*, 77-86.

[10] Wang, J., Han, J., Lu, Y., et., Tzvetkov, P. (2005). An efficient algorithm for mining top-k frequent closed itemsets, *IEEE Transaction on Knowledge and Data Engineering*, 17, 652–664.

[11] Ghezaiel, L. B., Latiri, C., et., Ahmed, M. B. (2012). Ontology Enrichment based on Generic Basis of Association Rules for Conceptual Document Indexing, *International Conference on Knowledge Engineering and Ontology Development*, 53-65.

[12] Paiva, L. (2015). Semantic relations extraction from unstructured information for domain ontologies enrichment.

[13] Maedche, A., et., Staab, S. (2000). Mining ontologies from Text, *In*: Proceedings of International Conference on Knowledge Engineering and Knowledg Management, 189–202.

[14] Maedche, A., et., Staab, S. (2001). Ontology learning for the semantic web, *IEEE Intelligent Systems*, 16 (12) 72–79.

[15] Paiva, L., Figueiras, P., et., Lima, C. (2014). Discovering Semantic Relations from Unstructured Data for Ontology Enrichment Asssociation rules based approach, 9th *Iberian Conference on Information Systems and Technologies (CISTI)*, 1-6.

[16] Fatemi, N., Poulin, F., Raileany, L. E., et., Smeaton, A. F. (2009). Using association rule mining to enrich semantic concepts for video retrieval, *International Conference on Knowledge Discovery and Information Retrieval*.

[17] d'Amato, C., Écrivain. (2014). *On extracting Rules for: enriching ontological knowledge bases, complementing heterogeneous sources of information, empowering the reasoning process*. [Performance]. Department of Computer Science University of Bari, Italy.

[18] Marinica, C., et., Guillet, F. (2010). Knowledge-based interactive postmining of association rules using ontologies,» *IEEE Transactions on Knowledge and Data Engineering*, 22, 784–797.

[19] Azevedo, P. J., et., Jorge, A. (2007). Comparing rulemeasures for predictive association rules, *Proceedings of the 18th European Conference on Machine Learning.*, 510–517, 2007.

[20] Idoudi, R., Hamrouni, K., et., Solaiman, B. (2014). Ontological Approach to Mammographic Knowledge Representation, *International Conference on Advanced Technologies*, 31-34.