Semantic Web 1 (0) 1–5
IOS Press

Machine Learning for the Semantic Web: Lessons Learnt and Next Research Directions

Claudia d'Amato
Department of Computer Science - University of Bari,
Via Orabona, 4 - 70126 Bari, Italy
E-mail: claudia.damato@uniba.it

Abstract. Machine Learning methods have been introduced in the Semantic Web for solving problems such as link and type prediction, ontology enrichment and completion (both at terminological and assertional level). Whilst initially mainly focussing on symbol-based solutions, recently numeric-based approaches have received major attention, motivated by the need to scale on the very large Web of Data. In this paper, the most representative proposals, belonging to the aforementioned categories are surveyed jointly with an analysis of their main peculiarities and drawbacks, afterwards the main envisioned research directions for further developing Machine Learning solutions for the Semantic Web are presented.

1. Introduction

The Semantic Web (SW) vision has been introduced with the goal of making the Web machine readable [1] by enriching resources with metadata, whose formal semantics is defined within OWL ontologies acting as shared vocabularies to be reused. Ontologies are also empowered with deductive reasoning capabilities which allow to derive knowledge that is implicitly encoded. While developing this vision, some limitations [2, 3] arose: ontology construction resulted in a time consuming task; furthermore, being strongly decoupled, ontologies and assertion can be out-of-sync, thus resulting incomplete, noisy and sometimes inconsistent with regard to the actual usage of the conceptual vocabulary in the assertions. These limitations became even more evident when pushing on Linked Data [4, 5] for enabling the actual creation of the Web of Data.

In this complex scenario multiple necessities emerged: reasoning at large scale; managing noise, inconsistencies and incompleteness in the Web of Data; (semi-)automatizing tasks such as ontology completion, enrichment (both at schema and assertional level) and link prediction; exploiting alternative forms of reasoning for complementing deductive reasoning. In order to fill some these gaps, machine learning (ML) methods have been proposed [6].

Problems such as query answering, instance retrieval and link prediction have been regarded as classification problems. Suitable machine learning methods, often inspired by symbol-based solutions in the *Inductive Logic Programming* (ILP) field, have been proposed [7–11]. Most of these solutions are able to cope with the expressive SW representations and the *Open World Assumption* (OWA) typically adopted, differently from the *Closed Wold Assumption* (CWA) usually assumed in the traditional ML settings.

Problems such as ontology refinement and enrichment at terminological level, e.g. assessing complex descriptions for a given concept name or assessing disjointness axioms, have been regarded as concept learning problems to be solved via supervised/unsupervised inductive learning methods for Description Logics [12] (DLs) representations [13–18].

Nowadays, the adoption of ML methods represents a major trend in several research fields such as computer vision, bioinformatics, image recognition, natural language processing and artificial intelligence. This is mostly due to the impressive scalability that recent ML methods, mostly grounded on numeric approaches (also called sub-symbolic) such as *embeddings* and *deep learning* [19], have shown. This phenomenon has also occurred in SW, where many of the recent ML-based solutions, e.g. for performing link and type pre-

1570-0844/0-1900/$35.00

dictions as well as data-intensive tasks by exploiting the Web of Data and the emerging Knowledge Graphs (KGs) as background knowledge, are mainly grounded on *embeddings*, motivated by scalability [20–22].

Nevertheless, the important gain, in terms of scalability, that ML methods for the SW are obtaining is penalizing: a) the possibility to have interpretable models as a result of a learning process; b) the ability to exploit deductive (and complementary forms) reasoning capabilities; c) the expressiveness of the SW representations and the compliance with the OWA.

In the following, the main problems and ML methods that have been developed in the SW are surveyed along with the two categories: symbol-based (Sect. 2) and numeric-based (Sect. 3), hence the fundamental peculiarities and issues are discussed. The main envisioned research directions that need to be pursued for developing ML methods for the SW are illustrated is Sect. 4. Conclusions are drawn in Sect. 5.

2. Symbol-based Methods for the Semantic Web

The first efforts in developing ML methods for the SW have been devoted to solve deductive reasoning tasks over ontologies under an inductive perspective. The motivation was given by the necessity of offering an alternative way to perform some forms of reasoning when deductive reasoning was not applicable, for instance because of inconsistencies within ontologies, but also for supplying a solution for reasoning in presence of incompleteness, that is when missing information with respect to a certain domain of reference is registered, and/or in presence of noise, that is when ontologies are consistent but the information therein is somehow wrong with respect to a reference domain, e.g. missing disjointness axioms, missing and/or wrong assertions. Successively, the incompleteness of knowledge bases, both at assertional and schema level, drove the development of ML methods trying to tackle this problem. The overall underlying idea consisted in exploiting the evidence coming from assertional knowledge for drawing plausible conclusions to be possibly represented with intensional models. In the following, the tasks that received major attention are reported jointly with the analysis of the main proposed solutions for them.

2.1. Instance Retrieval

One of the first problem that has been investigated is the *instance retrieval* problem, which amounts at assessing if an individual is an instance of a given concept. It has been regarded as a classification problem aiming at assessing the class-membership of an individual with respect to a query concept. Similarity-based methods, such as *K-Nearest Neighbor* and *Support Vector Machine*, have been developed since they are well known to be noise tolerant [7, 23, 24]. This required to cope with: 1) the OWA rather than the CWA generally adopted in ML; 2) the non-disjointness of the classes (since an individual can be instance of more that one concept at the same time) while, in the usual ML setting, classes are assumed to be disjoint; 3) the definition of new *similarity measures* and *kernel functions* for exploiting the expressiveness of semantic representations. Additionally, because of the OWA, new metrics for the evaluation of the classification results have been defined [7]. This is because, by using standard metrics as precision, recall and F-measure, new inductive results were deemed as mistakes whilst they could turn out to be correct inferences when judged by a knowledge engineer. The proposed methods experimentally proved their ability to perform inductive instance retrieval when compared to a standard deductive reasoner. They also proved their ability to induce new knowledge that was not logically derivable, but they did not result fully able to work at large scale.

Methods characterized by more interpretable models have been also defined [15, 25]. Inspired by the ILP literature for the induction of decision trees in clausal representation [26], a solution for inducing a *Terminological Decision Tree* (TDT) has been formalized [15]. A TDT is a tree structure, naturally compliant with the OWA, employing: a DL language for representing nodes; inference services as corresponding tests on the nodes. The tree-induction algorithm adopts a classical top-down divide-and-conquer strategy with the use of refinement operators for DL concept descriptions. Once a TDT is induced, similarly to logical decision trees, a definition for the target concept (namely the concept with respect to perform classification) can be drawn, by exploiting the nodes in the tree structure. This solution showed the interesting ability to provide an interpretable model, but it turned out slightly less effective then similarity-based classification methods.

2.2. Concept Learning for Ontology Enrichment

With the purpose of enriching ontologies at terminological level, methods for learning concept descriptions for a concept name have been proposed. The problem has been regarded as a supervised concept learning problem aiming at approximating an intensional DLs definition, given a set of individuals of an ontology acting as positive/negative training examples.

Various solutions, e.g. DL-FOIL [13] and CELOE [16] (part of the DL-LEARNER suite[1]), have been formalized. They are mostly grounded on a *separate-and-conquer* (sequential covering) strategy: a new concept description is built by specializing, via suitable *refinement operators*, a partial solution to correctly cover (i.e. decide a consistent classification for) as many training instances as possible. Whilst DL-FOIL works under OWA, CELOE works under CWA. Both of them may suffer of ending up in sub-optimal solutions. In order to overcome such issue, DL-FOCL [27], PARCEL [28] and SPACEL [29] have been proposed. DL-FOCL is an optimized version of DL-FOIL, implementing a base greedy covering learner. PARCEL combines top-down and bottom-up refinements in the search space. The learning problem is split into various sub-problems, according to a divide-and-conquer strategy, that are solved by running CELOE. Once the partial solutions are obtained, they are combined in a bottom-up fashion. SPACEL extends PARCEL with a symmetrical specialization of a concept description.

These solutions proved their ability to learn approximated concept descriptions for a target concept name but relatively small ontological knowledge bases have been considered for the experiments.

2.3. Knowledge Completion

Knowledge completion consists in finding new information at assertional level, that is facts, that is missing in a considered knowledge base. This task has become very popular with the development of KGs, that are well known to be incomplete, and it is also strongly related to the link prediction task (see Sect. 3).

One of the most well known system for knowledge completion of RDF knowledge bases is AMIE [10]. Inspired by the literature in *association rule mining* [30] and ILP methods for learning Horn clauses, AMIE aims to mine logic rules from RDF knowledge bases with the final goal of predicting new assertions. AMIE (and its optimized version AMIE+ [31]) currently represents the most scalable rule mining system for learning Horn rules on large RDF data collections and is also explicitly tailored to support the OWA. However, it does not exploit any form of deductive reasoning. A related rule mining system, similarly based on a level-wise generate and test strategy has been proposed in [32]. It aims to learn SWRL rules [33] from OWL ontologies while exploiting schema level information and deductive reasoning during the rule learning process. Both AMIE and the solution presented in [32] showed the ability to mine useful rules and to predict new assertional knowledge. However, because of the exploitation of the reasoning capabilities, the solution proposed in [32] showed reduced scalability.

2.4. Learning Disjointness Axioms

Disjointness axioms are essential for making explicit the negative knowledge about a domain, yet they are often overlooked during the modeling process (thus affecting the efficacy of reasoning services). To tackle this problem, automated methods for discovering these axioms from the data distribution have been devised.

A solution grounded on *association rule mining* [30] has been proposed in [17, 34]. It is based on studying the correlation between classes comparatively, namely *association rules*, *negative association rules* and *correlation coefficient*. Background knowledge and reasoning capabilities are used to a limited extent.

A different solution has been proposed in [18]. Here, moving from the assumption that two or more concepts may be mutually disjoint when the sets of their (known) instances do not *overlap*, the problem has been regarded as a clustering problem, aiming at finding partitions of similar individuals of the knowledge base, according to a *cohesion* criterion quantifying the degree of homogeneity of the individuals in an element of the partition. More specifically, the problem has been cast as a *conceptual clustering* problem, where the goal is both to find the best possible partitioning of the individuals and also to induce intensional definitions of the corresponding classes expressed in the standard representation languages. Emerging disjointness axioms are captured by the the employment of *terminological cluster trees*(TCTs) and by minimizing the risk of mutual overlap between concepts. Once the TCT is grown, groups of (disjoint) clusters located at sibling nodes identify concepts involved in candidate disjointness axioms to be derived. Unlike [17, 34], based on the statistical correlation between instances, the empirical evaluation of the method proposed in [18] showed its ability to discover disjointness axioms also involving complex concept descriptions thanks to the exploitation of the underlying ontology as a source of background knowledge.

[1] https://dl-learner.org/.

3. Numeric-based Methods for the Semantic Web

Whilst symbolic methods adopt symbols for representing entities and relationships of a domain and infer generalizations that provide new insights into the data and are ideally readily interpretable, numeric-based methods typically adopt feature vector (propositional) representations and cannot provide interpretable models but they usually result rather scalable [35].

The main problem that has been investigated in the SW context by adopting numeric solutions is *link prediction* which amounts to predict the existence (or the probability of correctness) of triples in (a portion of) the Web of Data. Data are considered in their graph representation and mostly RDF has been targeted as a representation language. Almost any reasoning is exploited and most expressive SW representation languages are basically discarded. The attention towards this problem is also grown due to the increasing of KGs, that are known to be often missing facts [36]. In this context, link prediction is also referred to as *knowledge graph completion*. Methods borrowed from the Statistical Relational Learning (SRL) [37] have been mostly developed. SRL main goal is the creation of statistical models for relational/graph-based data.

In the following the main classes of methods and the most representative solutions that have been developed targeting link prediction in the SW are analyzed.

3.1. Probabilistic Latent Variable Models

Probabilistic Latent Variable Models explains relations between entities by associating each resource to a set of intrinsic latent attributes (i.e. attributes not directly observable in the data) and conditions the probability distribution of the relations between two resources on their latent attributes. All relations are considered conditionally independent given the latent attributes. This allows the information to propagate through the network of interconnected latent variables.

One of the first numeric-based link prediction solution belonging to this category is the *Infinite Hidden Semantic Model* (IHSM) [38]. It formalizes a probabilistic latent variable that associates a latent class variable with each resource/node and makes use of constraints expressed in First Order Logic during the learning process. IHSM showed promising results but resulted limited in scaling on large SW data collection because of the complexity of the probabilistic inference and learning, which is intractable in general [39].

3.2. Embedding Models

With the goal of scaling on very large SW data collections, *embedding models* (also called energy-based models) have been investigated. Similarly to probabilistic latent variable models, in embedding models each resource/node is represented with a continuous embedding vector encoding its intrinsic latent features within the data collection. Models in this class do not necessarily rely on probabilistic inference for learning the optimal embedding vectors and this allows to avoid the issues related to the normalization of probability distributions, that may lead to intractable problems.

One of the first proposed solution belonging to this category is RESCAL [20], which implements graph embedding by computing a three-way factorization of an adjacency tensor that represents the multi-graph structure of the data collection. RESCAL resulted a powerful model also is was able to capture complex relational patterns over multiple hops in a graph, however, even if improving the scalability of IHSM, it did not result to be able to scale on very large graph-based data collection (e.g. the whole YAGO or DBPedia). The main limitation was represented by the parameter learning phase, which may take rather long for converging to optimal solutions.

However, since embedding models proved interesting ability to scale while maintaining comparative performance to probabilistic latent variable models in terms of predictive accuracy [40], with the goal of improving the model training phase employed by RESCAL, a solution exploiting adaptive learning rates during training has been proposed [21]. Specifically, an energy-based embedding model has been formalized, where entities and relations are embedded in continuous vector spaces and the probability of an RDF triple to encode a true statement is expressed in terms of energy of the triple, which is an unnormalized score that is inversely proportional to such a probability value. It is computed as a function of the embedding vectors of the subject, the predicate and the object of the triple. This solution experimentally showed improvements in terms of efficiency of the parameter learning process and more accurate results in a significantly lower number of iterations.

An important point that needs to be highlighted is that due to tackling RDF representation most of the considered data collections only contain positive (training) examples, since usually false facts are not encoded. As training a learning model in all-positive examples could be tricky because the model might easily over generalize, for obtaining negative examples two different approaches are generally adopted: either *perturbing* true/observed triples with the goal of generating plausible negative examples or making a *local-closed world assumption* (LCWA) in which the data collection is assumed as *locally* complete [41].

3.3. Vector Space Embeddings for Propositionalization

A complementary research direction is represented by works focussing on the exploitation of vector space embeddings for obtaining a propositional feature vector representation of RDF data collections. Specifically, inspired by the data mining (DM) literature on propositionalization [42], that is a collection of methods for transforming a relational data representation into a (numeric) propositional feature vector representation so that scalable propositional DM/ML methods can be applied, RDF2Vec [43] has been proposed. It formalizes a solution for learning latent numerical representations of entities in RDF graphs by adapting language modeling approaches. A two-steps approach is adopted: first the RDF graph is converted into a set of sequences of entities (for the purpose two different approaches using local information, that are graph walks and Weisfeiler-Lehman Subtree RDF graph kernels, are exploited); hence in the second step, the obtained sequences are used to train a neural language model estimating the likelihood of a sequence of entities appearing in a graph. The outcome of the the training process provides each entity in the graph represented as a vector of latent numerical features. DBpedia and Wikidata have been processed. In order to show that the obtained vector representation is task and algorithm independent, an experimental evaluation involving a number of classification and regression tasks has been performed.

An upgrade of RDF2Vec has been presented in [22]. Here, the proposed solution is grounded on the exploitation of global patterns, differently from RDF2Vec which exploits local patterns. None of the two solutions can cope with literals.

4. Machine Learning for the Semantic Web: Next Research Directions

In this section the envisioned most challenging research problems are illustrated. Hence additional ML settings and methods that could be usefully adopted for SW related issues are briefly discussed.

4.1. Research Problems

The need to cope with the fast growing of the Web of Data and the emerging very large KGs required the SW community to show its ability to manage such tremendous amount of data and knowledge, that it strongly contributed to create and grow up.

This mostly motivated the right attention towards numeric ML methods, particularly for providing scalable solutions to manage the inherent incompleteness of the Web of Data. Indeed, current symbolic methods are not actually comparable, in terms of scalability, to numeric-based solutions. This gain is not for free. It is obtained by giving up the expressive representation languages, such as OWL, that the SW community contributed to standardize with the goal of formalizing rich and expressive knowledge, but also by forgetting one of the most powerful characteristic of these languages, that is being empowered with deductive reasoning capabilities that allow to derive new knowledge. This means to loose knowledge that is already available. Indeed, as illustrated in Sect. 3, almost all numeric methods focus on RDF as a representation language and nearly no reasoning capabilities are exploited. Furthermore, differently from symbolic methods, numeric-based solutions lack of the ability to provide interpretable models (see Sect. 3), thus limiting the possibility to interpret and understand the motivations for the returned results. Additionally, tasks such as learning concept or disjointness axioms cannot be performed without symbol-based methods which can certainly benefit of very large amount of information to provide potentially more accurate results.

Integration of symbolic and numeric approaches: Research efforts need to be devoted towards ML solutions that, while keeping scalability, are able to target most expressive representations as well as to provide interpretable models. As a first step, the integration of numeric and symbolic approaches should be focused.

Some discussions in this direction have been developed in the Neural-Symbolic Learning and Reasoning community [44], which seeks to integrate principles from neural networks learning and logical reasoning. The main the conclusion has been that neural-symbolic integration appears particularly suitable to applications characterized by the joint availability of large amounts of (heterogeneous) data and knowledge descriptions, which is actually the case of the Web of Data. A set of key challenges and opportunities have been also outlined [44], such as: how representing expressive logics within neural networks, how neural networks should reason with variables, or how to extract symbolic representation from trained neural networks.

Some preliminary results for some these challenges have been recently provided. In [45], a scalable tensor-based factorization model, called SimplE, has been formalized. Besides scalability, its main peculiarity is the ability to learn interpretable embeddings incorporating logical rules through weight tying. In [46], ideas for extracting propositional rules from trained neural networks under a SW background knowledge are illustrated, showing that the exploitation of a background knowledge allows: to reduce the extracted rule set; to reproduce the input-output function of the trained neural network. A conceptual sketch for explaining artificial neural networks classification behavior in a non-propositional setting while using SW background knowledges have been proposed in [47]. These initial results, show the feasibility of this research direction while remarking the importance of pursuing this goal.

Providing Explanations: The problem of explaining artificial neural networks classification behavior, tackled in [47], sheds the light on another important issue argued in [48], that is the necessity to provide explanations for results supplied by ML methods, particularly when they come from very large sources of knowledge, e.g. results for a link prediction problem. The solution depicted in [47] is in agreement with the idea of exploiting symbol-based interpretable models to explain conclusions [35, 49]. Nevertheless, interpretable models describe *how* solutions are obtained but not *why* they are obtained. As argued in [44, 50], providing an explanation means to supply a line of reasoning, illustrating the decision making process of a model whilst using human understandable features. Following this direction, a solution providing human-centric transfer learning explanation has been proposed [51]. It takes advantage of ontologies (DBPedia is used) and reasoning capabilities to infer different kinds of human understandable explanatory evidence.

Hence, providing an explanation means to open the box of the reasoning process and and makes it understandable. In a complex setting such as the Web of Data, where knowledge may be the result of an automatic information acquisition and integration process from different sources, thus potentially noisy and with conflicting information, multiple reasoning paradigms may be required e.g. deductive reasoning (when rules and theory are available), inductive reasoning (for building models from the available knowledge), abductive reasoning (for filling in partial models coping with incomplete theory), commonsense reasoning etc. Large research efforts have been devoted to study each paradigm, however in the considered complex scenario, multiple paradigms could be needed at the same time. This may require the formalization of a unifying reasoning framework.

Capturing Context and Evolution: Some acquired knowledge may also evolve over the time, that is it can be valid only for a certain time period, or it may be context dependent [52]. Capturing these phenomenon may be fundamental and unsupervised and pattern mining methods would be useful for the purpose. Some preliminary research on capturing knowledge evolution by conceptual clustering methods has been presented [53] showing the feasibility of the approach but highlighting an existing limitation given by the lack of gold standards for validating the results.

4.2. Additional Machine Learning Settings

As for the settings to be exploited, multiple research directions still need to be investigated. Several problems such as instance retrieval but also link prediction and assertional knowledge completion have been solved by casting them as classification tasks. However, as discussed in Sect. 2, when assessing the concept membership for an individual, it may result instance of more than one concept at the same time. As such a more suitable way to regard the problem is as *multi-label classification* task [54], where multiple labels (concepts in the specific case) may be assigned to each instance. Some preliminary research has been presented in [55], focussing on type prediction in RDF data collections where limited information from the available background knowledge is considered.

Multiple-instance learning (MIL) [56] is also a setting that would need investigation. It deals with problems with incomplete knowledge of labels in training sets, as it happens in SW knowledge bases due to OWA. MIL is a type of supervised learning where training instances are not individually labeled, they are collected in sets of labeled bags, each one containing many instances. From a collection of labeled bags, the learner tries to either (i) induce a concept that will label individual instances correctly or (ii) learn how to label bags without inducing the concept. It may be fruitfully exploited for discovering correlations among resources and/or emerging concepts.

Other settings that would be useful for coping with the large number of unlabelled instances are *semi-supervised learning* (SSL) [57] and *learning from imbalanced data*. SSL makes use of both labeled and unlabeled instances, during the learning process, for surpassing the classification performance that could be obtained by discarding the unlabeled data (as it would happen in a supervised learning setting). Very few research efforts have been made in this direction. Some initial results have been presented in [58], were a link prediction problem is solved in a transductive learning framework. In learning from unbalanced data [59, 60], that is data collections where the labels distribution is not uniform, sampling techniques are usually adopted in order to create a balanced dataset to be successively used for the learning task. *Ensemble methods*, consisting in using multiple learning algorithms to obtain better predictive performance, could be fruitfully adopted, as illustrated in [25, 61] where respectively a *boosting* [61] and *bagging* technique is employed.

As a last point, considering the increasing volume of the Web of Data, *online* and *incremental learning*, which input data is continuously used to further train and extend the learned model, would be naturally investigated. For the best of knowledge, no research efforts have been made in this direction.

5. Conclusions

In this paper, the progresses that have been made in SW by exploiting ML methods have been surveyed. Specifically symbol-based and numeric-based solutions have been analyzed highlighting their main peculiarities and drawback. Hence the main envisioned research directions have been drawn.

References

[1] Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. Scientific American (2001)

[2] Siorpaes, K., Hepp, M.: Ontogame: Towards overcoming the incentive bottleneck in ontology building. In Meersman, R., et al., eds.: OTM 2007 Workshop Proc. Volume 4806 of LNCS. Springer (2007)

[3] Hoekstra, R.: The knowledge reengineering bottleneck. Semantic Web Journal **1** (2010) 111–115

[4] Shadbolt, N., Berners-Lee, T., Hall, W.: The semantic web revisited. IEEE Intelligent Systems **21** (2006) 96–101

[5] Bizer, C., Heath, T., Berners-Lee, T.: Linked data – the story so far. Int. J. on Sem. Web and Inf. Syst. **5** (2009) 1–22

[6] d'Amato, C., Fanizzi, N., Esposito, F.: Inductive learning for the semantic web: What does it buy? Sem. Web **1** (2010) 53–59

[7] d'Amato, C., Fanizzi, N., Esposito, F.: Query answering and ontology population: An inductive approach. In Bechhofer, S., et al., eds.: ESWC 2008 Proceedings. Volume 5021 of LNCS., Springer (2008) 288–302

[8] Józefowska, J., Lawrynowicz, A., Lukaszewski, T.: The role of semantics in mining frequent patterns from knowledge bases in description logics with rules. TPLP **10** (2010) 251–289

[9] Lehmann, J., Bühmann, L.: ORE - a tool for repairing and enriching knowledge bases. In: Proceedings of ISWC 2010. Volume 6497 of LNCS., Springer (2010) 177–193

[10] Galárraga, L., Teflioudi, C., Hose, K., Suchanek, F.M.: AMIE: association rule mining under incomplete evidence in ontological knowledge bases. In Schwabe, D., et al., eds.: WWW'13 Proc., ACM (2013) 413–422

[11] Tiddi, I., d'Aquin, M., Motta, E.: Dedalo: Looking for clusters explanations in a labyrinth of linked data. In Presutti, V., et al., eds.: ESWC Proceedings. Volume 8465 of LNCS., Springer (2014) 333–348

[12] Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F., eds.: Description Logic Handbook, 2nd edition. Cambridge University Press (2010)

[13] Fanizzi, N., d'Amato, C., Esposito, F.: DL-FOIL concept learning in description logics. In: ILP. Volume 5194 of Lecture Notes in Computer Science., Springer (2008) 107–121

[14] Fanizzi, N., d'Amato, C., Esposito, F.: Metric-based stochastic conceptual clustering for ontologies. Inf. Syst. **34** (2009) 792–806

[15] Fanizzi, N., d'Amato, C., Esposito, F.: Induction of concepts in web ontologies through terminological decision trees. In: ECML/PKDD (1). Volume 6321 of LNCS., Springer (2010) 442–457

[16] Lehmann, J., Auer, S., Bühmann, L., Tramp, S.: Class expression learning for ontology engineering. J. Web Semant. **9** (2011) 71–81

[17] Völker, J., Fleischhacker, D., Stuckenschmidt, H.: Automatic acquisition of class disjointness. Journal of Web Semantics **35** (2015) 124–139

[18] Rizzo, G., d'Amato, C., Fanizzi, N., Esposito, F.: Terminological cluster trees for disjointness axiom discovery. In Blomqvist, E., et al., eds.: Proceedings of ESWC 2017, Part I. Volume 10249 of LNCS., Springer (2017) 184–201

[19] Deng, L., Yu, D., eds.: Deep Learning: Methods and Applications. Now Pub (2014)

[20] Nickel, M., Tresp, V., Kriegel, H.: A three-way model for collective learning on multi-relational data. In: Proc. of ICML 2011. Omnipress (2011) 809–816

[21] Minervini, P., d'Amato, C., Fanizzi, N.: Efficient energy-based embedding models for link prediction in knowledge graphs. J. Intell. Inf. Syst. **47** (2016) 91–109

[22] Cochez, M., Ristoski, P., Ponzetto, S., Paulheim, H.: Global RDF vector space embeddings. In: ISWC (1) Proc. Volume 10587 of LNCS. Springer (2017) 190–207

[23] Rettinger, A., Lösch, U., Tresp, V., d'Amato, C., Fanizzi, N.: Mining the semantic web - statistical learning for next generation knowledge bases. Data Minining and Knowledge Discovery **24** (2012) 613–662

[24] Bloehdorn, S., Sure, Y.: Kernel methods for mining instance data in ontologies. In: ISWC/ASWC. Volume 4825 of LNCS., Springer (2007) 58–71

[25] Rizzo, G., Fanizzi, N., d'Amato, C., Esposito, F.: Approximate classification with web ontologies through evidential terminological trees and forests. Int. J. Approx. Reasoning **92** (2018) 340–362

[26] Blockeel, H., De Raedt, L.: Top-down induction of first-order logical decision trees. Artif. Intelligence **101** (1998) 285–297

[27] Rizzo, G., Fanizzi, N., d'Amato, C., Esposito, F.: A framework for tackling myopia in concept learning on the web of data. In: EKAW. Volume 11313 of LNCS., Springer (2018) 338–354

[28] Tran, A., Dietrich, J., Guesgen, H., Marsland, S.: An approach to parallel class expression learning. In Bikakis, A., Giurca, A., eds.: RuleML 2012. Volume 7438 of LNCS. Springer (2012) 302–316

[29] Tran, A., Dietrich, J., Guesgen, H., Marsland, S.: Parallel symmetric class expression learning. J. of Machine Learning Research **18** (2017) 1–34

[30] Agrawal, R., Imieliński, T., Swami, A.: Mining association rules between sets of items in large databases. In: SIGMOD. (1993) 207–216

[31] Galárraga, L., Teflioudi, C., Hose, K., Suchanek, F.M.: Fast rule mining in ontological knowledge bases with AMIE+. VLDB J. **24** (2015) 707–730

[32] d'Amato, C., Tettamanzi, A., Tran, D.M.: Evolutionary discovery of multi-relational association rules from ontological knowledge bases. In: EKAW 2016. Volume 10024 of LNCS. (2016) 113–128

[33] Horrocks, I., Patel-Schneider, P., Boley, H., Tabet, S., Grosof, B., Dean., M.: SWRL: A semantic web rule language combining owl and ruleml. W3C (2004)

[34] Völker, J., Niepert, M.: Statistical schema induction. In Antoniou, G., et al., eds.: Proceedings of ESWC 2011. Volume 6643 of LNCS. Springer (2011) 124–138

[35] Luger, G.F., ed.: Artificial Intelligence: Structures and Strategies for Complex Problem Solving. Addison Wesley (2005)

[36] West, R., Gabrilovich, E., Murphy, K., Sun, S., Gupta, R., Lin, D.: Knowledge base completion via search-based question answering. In: Proc. of WWW 2014. ACM (2014) 515–526

[37] Getoor, L., Taskar, B., eds.: Introduction to Statistical Relational Learning. MIT Press (2007)

[38] Rettinger, A., Nickles, M., Tresp, V.: Statistical relational learning with formal ontologies. In Buntine, W., et al., eds.: Proc. of ECML/PKDD 2009, Part II. Volume 5782 of LNCS. Springer (2009) 286–301

[39] Koller, D., Friedman, N., eds.: Probabilistic graphical models: principles and techniques. MIT Press (2009)

[40] Bordes, A., Gabrilovich, E.: Constructing and mining web-scale knowledge graphs. In Gangemi, A., at al., eds.: Proc. of WWW 2015 - companion volume. ACM (2015)

[41] Nickel, M., Murphy, K., Tresp, V., Gabrilovich, E.: A review of relational machine learning for knowledge graphs. Proceedings of the IEEE **104** (2016) 11–33

[42] Kramer, S., Lavrač, N., Flach, P.: Propositionalization approaches to relational data mining. In Džeroski, S., Lavraž, N., eds.: Relational Data Mining. Springer (2001) 262–291

[43] Ristoski, P., Paulheim, H.: Rdf2vec: RDF graph embeddings for data mining. In: International Semantic Web Conference (1). Volume 9981 of LNCS. Springer (2016) 498–514

[44] d'Avila Garcez, A., Besold, T., De Raedt, L., Földiák, P., Hitzler, P., Icard, T., Kühnberger, K., Lamb, L., Miikkulainen, R., Silver, D.: Neural-symbolic learning and reasoning: Contributions and challenges. In: AAAI Spring Symposia. AAAI Press (2015)

[45] Kazemi, S., Poole, D.: Simple embedding for link prediction in knowledge graphs. In Bengio, S., et al., eds.: NeurIPS 2018 Proc. (2018) 4289–4300

[46] Labaf, M., Hitzler, P., Evans, A.: Propositional rule extraction from neural networks under background knowledge. In: NeSy. Volume 2003 of CEUR Work. Proc., CEUR-WS.org (2017)

[47] Sarker, M., Xie, N., Doran, D., Raymer, M., Hitzler, P.: Explaining trained neural networks with semantic web technologies: First steps. In: NeSy. Volume 2003 of CEUR Workshop Proceedings., CEUR-WS.org (2017)

[48] d'Amato, C.: Logic and learning: Can we provide explanations in the current knowledge lake? In Bonatti, P., et al., eds.: Knowledge Graphs: New Directions for Knowledge Representation on the Semantic Web (Dagstuhl Seminar 18371). Volume 8 of Dagstuhl Reports. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik (2019) 37–38

[49] Raedt, L.D., ed.: Logical and Relational Learning: From ILP to MRDM (Cognitive Technologies). Springer-Verlag (2008)

[50] Doran, D., Schulz, S., Besold, T.: What does explainable AI really mean? A new conceptualization of perspectives. In Besold, T., Kutz, O., eds.: Proc. of the Comprehensibility and Explanation in AI and ML Workshop. Volume 2071 of CEUR Work. Proc. CEUR-WS.org (2018)

[51] Chen, J., Lécué, F., Pan, J., Horrocks, I., Chen, H.: Knowledge-based transfer learning explanation. In Thielscher, M., et al., eds.: KR 2018 Proc., AAAI Press (2018) 349–358

[52] d'Amato, C., Kirrane, S., Bonatti, P., Rudolph, S., Krötzsch, M., van Erp, M., Zimmermann, A.: Foundations. In Bonatti, P., et al., eds.: Knowledge Graphs: New Directions for Knowledge Representation on the Semantic Web (Dagstuhl Seminar 18371). Volume 8 of Dagstuhl Reports. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik (2019) 79–88

[53] Fanizzi, N., d'Amato, C., Esposito, F.: Conceptual clustering and its application to concept drift and novelty detection. In Bechhofer, S., et al., eds.: ESWC 2008 Proc. Volume 5021 of LNCS., Springer (2008) 318–332

[54] Zhou, Z., Zhang, M.: Multi-label learning. In Sammut, C., Geoffrey, W., eds.: Encyclopedia of Machine Learning and Data Mining. Springer US (2017) 875–881

[55] Melo, A., Völker, J., Paulheim, H.: Type prediction in noisy RDF knowledge bases using hierarchical multilabel classification with graph and latent features. International Journal on Artificial Intelligence Tools **26** (2017) 1–32

[56] Carbonneau, M., Cheplygina, V., Granger, E., Gagnon, G.: Multiple instance learning. Pattern Recogn. **77** (2018)

[57] Chapelle, O., Schölkopf, B., Zien, A., eds.: Semi-supervised learning. Cambridge MIT Press (2006)

[58] Minervini, P., Tresp, V., d'Amato, C., Fanizzi, N.: Adaptive knowledge propagation in web ontologies. TWEB **12** (2018) 2:1–2:28

[59] Guo, H., Herna, V.L.: Learning from imbalanced data sets with boosting and data generation: the databoost-im approach. SIGKDD Explor. Newsl. **6** (2004) 30–39

[60] Liu, X., Wu, J., Zhou, Z.: Exploratory under-sampling for class-imbalance learning. In: ICDM 2006 Proc., IEEE Computer Society (2006) 965–969

[61] Fanizzi, N., Rizzo, G., d'Amato, C.: Boosting dl concept learners. In: ESWC 2019 Proc. LNCS, Springer (2019)