Metadata of the chapter that will be visualized in SpringerLink

Book Title	Reasoning Web. Declarative Artificial Intelligence
Series Title	
Chapter Title	Mining the Semantic Web with Machine Learning: Main Issues that Need to Be Known
Copyright Year	2022
Copyright HolderName	Springer Nature Switzerland AG

Corresponding Author	Family Name	**d'Amato**
	Particle	
	Given Name	**Claudia**
	Prefix	
	Suffix	
	Role	
	Division	
	Organization	Dipartimento di Informatica – Università degli Studi di Bari Aldo Moro
	Address	Bari, Italy
	Email	claudia.damato@uniba.it

Abstract

The Semantic Web (SW) is characterized by the availability of a vast amount of semantically annotated data collections. Annotations are provided by exploiting ontologies acting as shared vocabularies. Additionally ontologies are endowed with deductive reasoning capabilities which allow to make explicit knowledge that is formalized implicitly. Along the years a large number of data collections have been developed and interconnected, as testified by the Linked Open Data Cloud. Currently, seminal examples are represented by the numerous Knowledge Graphs (KGs) that have been built, either as enterprise KGs or open KGs, that are freely available. All of them are characterized by very large data volumes, but also incompleteness and noise. These characteristics have made the exploitation of deductive reasoning services less feasible from a practical viewpoint, opening up to alternative solutions, grounded on Machine Learning (ML), for mining knowledge from the vast amount of information available. Actually, ML methods have been exploited in the SW for solving several problems such as link and type prediction, ontology enrichment and completion (both at terminological and assertional level), and concept leaning. Whilst initially symbol-based solutions have been mostly targeted, recently numeric-based approaches are receiving major attention because of the need to scale on the very large data volumes. Nevertheless, data collections in the SW have peculiarities that can hardly be found in other fields. As such the application of ML methods for solving the targeted problems is not straightforward. This paper extends [20], by surveying the most representative symbol-based and numeric-based solutions and related problems, with a special focus on the main issues that need to be considered and solved when ML methods are adopted in the SW field as well as by analyzing the main peculiarities and drawbacks for each solution.

Keywords (separated by '-')	Semantic Web - Machine learning - Symbol-based methods - Numeric-based methods

Mining the Semantic Web with Machine Learning: Main Issues that Need to Be Known

Claudia d'Amato(✉)

Dipartimento di Informatica – Università degli Studi di Bari Aldo Moro, Bari, Italy
claudia.damato@uniba.it

Abstract. The Semantic Web (SW) is characterized by the availability of a vast amount of semantically annotated data collections. Annotations are provided by exploiting ontologies acting as shared vocabularies. Additionally ontologies are endowed with deductive reasoning capabilities which allow to make explicit knowledge that is formalized implicitly. Along the years a large number of data collections have been developed and interconnected, as testified by the Linked Open Data Cloud. Currently, seminal examples are represented by the numerous Knowledge Graphs (KGs) that have been built, either as enterprise KGs or open KGs, that are freely available. All of them are characterized by very large data volumes, but also incompleteness and noise. These characteristics have made the exploitation of deductive reasoning services less feasible from a practical viewpoint, opening up to alternative solutions, grounded on Machine Learning (ML), for mining knowledge from the vast amount of information available. Actually, ML methods have been exploited in the SW for solving several problems such as link and type prediction, ontology enrichment and completion (both at terminological and assertional level), and concept leaning. Whilst initially symbol-based solutions have been mostly targeted, recently numeric-based approaches are receiving major attention because of the need to scale on the very large data volumes. Nevertheless, data collections in the SW have peculiarities that can hardly be found in other fields. As such the application of ML methods for solving the targeted problems is not straightforward. This paper extends [20], by surveying the most representative symbol-based and numeric-based solutions and related problems, with a special focus on the main issues that need to be considered and solved when ML methods are adopted in the SW field as well as by analyzing the main peculiarities and drawbacks for each solution. AQ1

Keywords: Semantic Web · Machine learning · Symbol-based methods · Numeric-based methods

1 Introduction

The Semantic Web (SW) vision has been introduced with the goal of making the Web machine readable [5], by enriching resources with metadata whose formal semantics is defined in OWL[1] ontologies acting as shared vocabularies to be reused. Ontologies are

[1] https://www.w3.org/OWL/.

M. Šimkus and I. Varzinczak (Eds.): Reasoning Web 2021, LNCS 13100, pp. 1–18, 2022.
https://doi.org/10.1007/978-3-030-95481-9_4

also empowered with deductive reasoning capabilities which allow for deriving knowledge that is implicitly encoded. While developing this vision, some limitations [35,67] arose: ontology construction resulted in a time consuming task; being strongly decoupled, ontologies and assertions can be out-of-sync, thus resulting in incomplete, noisy and sometimes inconsistent ontologies due to the actual usage of the conceptual vocabulary in the assertions. These limitations became even more evident when pushing on Linked Data [6,66] for enabling the actual creation of the Web of Data and nowadays with the progressive growth of Knowledge Graphs [36]. As a consequence, multiple necessities emerged: reasoning at large scale; managing noise, inconsistencies and incompleteness in the data collections; (semi-)automatizing tasks such as ontology completion, enrichment (both at schema and assertional level), link prediction; exploiting alternative forms of reasoning complementing the deductive approach.

In order to fill some of these gaps, machine learning (ML) methods have been proposed [17]. 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 (aiming at inducing a hypothesised logic program from background knowledge and a collection of examples), have been proposed [16,28,40,46,69]. Most of them are able to cope with the expressive SW representations and the *Open World Assumption* (OWA) typically adopted, differently from the *Closed Wold Assumption* (CWA) that is usually assumed in the traditional ML settings. Problems such as ontology refinement and enrichment at terminology level, e.g. assessing disjointness axioms or complex descriptions for a given concept name, have been regarded as concept learning problems to be solved via supervised/unsupervised inductive learning methods for Description Logic [4] (DLs) representations [24–26,45,62,73].

Nowadays, numeric-based (also called sub-symbolic) ML methods, such as *embeddings* [14,51,56] and *deep learning* [21], are receiving major attention because of their impressive ability to scale when applied to very large data collections. Mostly KG refinement tasks, such as link/type predictions and triple classifications are targeted, with the goal of improving/limiting incompleteness in KGs. Nevertheless, the important gain, in terms of scalability, that numeric-based 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 of) reasoning capabilities; c) the expressiveness of the SW representations to be considered 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 symbol-based (Sect. 2) and numeric-based (Sect. 3) categories, hence the fundamental peculiarities and issues are discussed. Afterwards, considerations concerning the need for solutions that are able to provide human understandable explanations and, towards this direction, to come up with a unified framework integrating both numeric and symbol-based solutions, are reported in 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. This was motivated 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, e.g. missing disjointness axioms), 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 and/or wrong (derivation of) assertions). Particularly, the incompleteness of knowledge bases, both at assertional and schema level, drove the development of ML methods trying to specifically tackle this problem. The overall 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 solutions for them.

2.1 Instance Retrieval

One of the first problems that has been investigated is the *instance retrieval* problem, which amounts to 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 [8,16,59]. 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 SW representations. Additionally, because of the OWA, new metrics for the evaluation of the classification results have been defined [16]. This is because, by using standard metrics such 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 new metrics do not have a direct mapping to the sets of true/false positives/negatives, rather because of the OWA, they consider the cases of unknown/unlabeled results. Particularly, *match rate*, *omission error*, *commission error* and *induction rate* have been proposed. The match rate measures the rate of classification results in agreement with the labels (provided by the use of a standard deductive reasoner). The omission rate measures the cases in which the inductive classifier was not able to provide results, due to the abundance of unlabeled instances because of the OWA, whilst actual labels were available. The commission error measures the cases in which the classifier provided results opposite to the true labels (e.g. an individual being instance of the negated query concept). The induction rate measures the cases in which the classifier was able to provide a label whilst it was not available due to OWA. The proposed solutions experimentally proved their ability to perform inductive instance retrieval when compared to a standard deductive reasoner. Additionally, they also proved their ability to induce new knowledge that was not logically derivable[2]. Nevertheless, they were not fully able to work at large scale.

Methods characterized by more interpretable models have also been defined [26, 63]. Inspired by the ILP literature on the induction of decision trees in clausal representation [7], a solution for inducing a *Terminological Decision Tree* (TDT) has been formalized [26]. A TDT is a tree structure, naturally compliant with the OWA, employing: a DL language for representing nodes and 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 which classification is to be performed) 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.

Nevertheless, when assessing the concept membership for an individual, as recalled above, 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 [77], where multiple labels (concepts in the specific case) may be assigned to each instance. Some preliminary research has been presented in [50], focussing on type prediction in RDF data collections where limited information from the available background knowledge is considered. *Multiple-instance learning* (MIL) [11] is also a setting that would need investigation. It deals with the problem of incomplete knowledge concerning 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. 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) [12] 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 [52], where a link prediction problem is solved in a transductive learning framework. In learning from unbalanced data [32,48], 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 [27,63] where respectively a *boosting* [27] and *bagging* [63] technique is employed.

[2] The induced knowledge should be validated by ontology engineerings for the possible further enrichment of ontologies.

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 [24] and CELOE [45] (part of the DL-LEARNER suite[3]), 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 from ending up in sub-optimal solutions. In order to overcome such issue, DL-FOCL [64], PARCEL [70] and SPACEL [71] 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 are 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 systems for knowledge completion of RDF knowledge bases is AMIE [28]. Inspired by the literature in *association rule mining* [2] 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+ [29]) 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 [19]. It aims to learn SWRL rules [37] from OWL ontologies while exploiting schema level information and deductive reasoning during the rule learning process. Both AMIE and the solution presented in [19] showed the ability to mine useful rules and to predict new assertional knowledge. The solution proposed in [19] showed reduced scalability due to the exploitation of the reasoning capabilities.

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

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* [2] has been proposed in [72,73]. 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 [62] where, 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. 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 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 [72,73], based on the statistical correlation between instances, the empirical evaluation of [62] showed its ability to discover disjointness axioms also involving complex concept descriptions, thanks to the exploitation of the underlying ontology as background knowledge.

2.5 Capturing Ontology Evolutions

Some acquired knowledge may also evolve over time. For instance, given an ontology, due to the insertion of new individuals and assertions, new concept formations may emerge over time but lacking of intentional definitions (*novelty detection* [68]). Similarly, existing concepts, defined by their intention, may actually evolve towards more general or more specific concepts when looking at they extensions that again may evolve over time (*concept drift* [75]).

Capturing these phenomenon may be fundamental and unsupervised as well as pattern mining methods would be useful for the purpose. Some preliminary research on capturing knowledge evolution by exploiting conceptual clustering methods has been presented [23,25]. Particularly, a two step approach is proposed. As a first step, suitable distance-based and semantically enhanced clustering method are exploited in oder to spot cases of concept that are evolving or novel concepts which are emerging based on the elicited clusters. Afterwards, concept learning algorithms for DL representations (see Sect. 2.2) are used to produce new concepts based on a group of examples (i.e. individuals in a cluster) and counterexamples (individuals in disjoint clusters).

The proposed solutions proved the feasibility of the overall approach by showing the ability to capture new and evolving concepts but also highlighted a main limitation given by the lack of gold standards for validating the results.

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 are usually rather scalable [49].

The problem that has been mainly 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, mostly the RDF representation language has been targeted and almost no reasoning is exploited; most expressive SW languages are basically discarded. The attention towards this problem is also grown due to the increasing availability of KGs, that are known to be often missing facts [74]. In the KG context, link prediction is also referred to as *knowledge graph completion*. Methods borrowed from the Statistical Relational Learning (SRL) [31] (having as main goal the creation of statistical models for relational/graph-based data) have been mostly developed. In the following the main classes of methods and solutions targeting link prediction in the SW are analyzed.

3.1 Probabilistic Latent Variable Models

Probabilistic Latent Variable Models explain 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) [60]. 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 in a limited scalability on large SW data collection because of the complexity of the probabilistic inference and learning, which is intractable in general [42].

3.2 Embedding Models

With the goal of scaling on very large SW data collections, *embedding 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.

Particularly, KG embedding methods have received considerable attention. They typically map entities and relations forming complex graph structures to simpler representations (feature-vectors) and aim at learning prediction functions to be exploited for tasks such as link prediction and triple classification. The scalability purpose motivated the interest delved towards these models [10] which have been shown to ensure good performances on very large KGs. Specifically, KG embedding methods aim at converting the data graph into an optimal low-dimensional space in which *graph structural information* and *graph properties* are preserved as much as possible [10,39]. The low-dimensional spaces enable computationally efficient solutions that scale better with the KG dimensions. Graph embedding methods may differ in their main building blocks: the *representation space* (e.g. point-wise, complex, discrete, Gaussian, manifold), the *encoding model* (e.g. linear, factorization, neural models) and the *scoring function* (that can be based on distance, energy, semantic matching or other criteria) [39]. In any case, the objective consists in learning embeddings such that the score of a valid (positive) triple is lower than the score of an invalid triple standing for a sort of negative examples.

One of the first solutions belonging to this category is RESCAL [56], 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 in a powerful model, it was also able to capture complex relational patterns over multiple hops in a graph, however, even if improving the scalability of IHSM, it was not able to scale on very large graph-based data collections (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. With the goal of improving the model training phase employed by RESCAL, a solution exploiting adaptive learning rates during training has been proposed [51]. 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.

Nevertheless, the very first embedding model that registered very high scalability performances has been TransE [9]. It introduces a very simple but effective and efficient model: each entity is represented by an embedding vector and each predicate is represented by a (vector) *translation operation*. The score of a triple is given by the similarity (negative L_1 or L_2 distance) of the translated subject embedding to the object embedding. The optimal embedding and translation vectors for predicates are learned jointly. The method relies on a *stochastic optimization process*, that iteratively updates the distributed representations by increasing the score of the positive triples i.e. the observed triples, while lowering the score of unobserved triples standing as negative examples. The embedding of all entities and predicates in the KG is learned by minimizing a *margin-based ranking loss*.

Despite the scalability of TransE, it resulted limited in representing properly various types of properties such as *reflexive* ones, and 1-to-N, N-to-1 and N-to-N relations. To tackle this limitation, while keeping the ability to scale on very large KGs, moving from TRANSE, a large family of models has been developed. Among others, TRANSR [47] has been proposed as a more suitable model to handle non 1-to-1 relations. It adopts a score function that preliminarily projects entities and relations to the different spaces and successively they are put together through a suitable projection matrix. The main variation introduced by the new model regards the way the entities are projected in the vector space of the relations, which increases the complexity without compromising the overall scalability.

An important point that needs to be highlighted is that, due to tackling RDF representations, 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 *corrupting* 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 [55]. In both cases, wrong negative information may be generated and thus used when training and learning the embedding models; hence alternative solutions are currently investigated [3]. Even more so, existing embedding models do not make use of the additional semantic information that may be coded when more expressive representation languages are adopted. Indeed the need for *semantic embedding methods* has been argued [20,38,57].

3.3 Semantically Enriched Embedding Models

Recently, semantically empowered embedding models, particularly targeting KG refinement tasks, have been investigated [18,38,53] and various approaches have been proposed that leverage different specific forms of prior knowledge to learn better representations.

In [33] a KG embedding method considering also logical rules has been proposed, where triples in the KG and rules are represented in a unified framework. Specifically, triples are represented as atomic formulae while rules are represented as more complex formulae modeled by t-norm fuzzy logics admitting as antecedent single atoms or conjunctions of atoms with variables as subjects and objects. A common loss over both representation is defined which is minimized to learn the embeddings. This proposal resulted in a novel solution but the specific form of prior knowledge that has to be available for the KG constitutes its main drawback. A similar drawback also applies to the model proposed in [54], where a solution based on adversarial training is formalized, exploiting Datalog clauses to encode assumptions which are used to regularize neural link predictors. An inconsistency loss is derived that measures the degree of violation of such assumptions on a set of adversarial examples. Training is defined as a minimax problem, in which the models are trained by minimizing the inconsistency loss on the adversarial examples jointly with a supervised loss. Nevertheless, in [1] the limitations of the current embedding models have been identified: theoretical inexpressiveness, lack of support for inference patterns, higher-arity relations, and logical rule incorporation.

Complementary solutions, besides exploiting the graph structural information and properties, focused on exploiting also the additional knowledge available, when rich representation languages as RDFS and OWL are employed, that is when no specific addtional formalisms are required for representing additional prior knowledge. Particularly, [53] has proven the effectiveness of combinations of embedding methods and strategies relying on reasoning services for the injection of *Background Knowledge* (BK) to enhance the performance of a specific predictive model. Following this line, TRANSOWL, aiming at injecting BK particularly during the learning process, and its upgraded version TRANSROWL, where a newly defined and more suitable loss function and scoring function are also exploited, have been proposed [18]. The main focus is on the application of this idea to enhance well-known basic scalable models, namely TRANSE [9] and TRANSR [47][4], even if, in principle, the proposed approach could be applied to more complex embedding methods, with an additional formalization. The proposed solutions can take advantage of an informed corruption process that leverages on reasoning capabilities, while limiting the amount of false negatives that a less informed random corruption process may cause.

In TRANSOWL the original TRANSE setting is maintained while resorting to reasoning with schema axioms to derive further triples to be considered for training and that are generated consistently with the semantics of the properties. Particularly, for each considered axiom, TRANSOWL defines, on the score function, specific constraints that guide the way embedding vectors are learned. It extends the approach in [53], formalizing a model characterized by two main components devoted to inject BK in the embedding-based model during the training phase: 1) *Reasoning:* It is used for generating corrupted triples that can certainly represent negative instances, thus avoiding false negatives, for a more effective model training. Moreover, false positives can be detected and avoided. Specifically, using a reasoner[5] it is possible to generate corrupted triples exploiting the available axioms specified in RDFS and OWL. The following axioms are considered: domain, range, disjointWith, functionalProperty; 2) *BK Injection:* A set of different axioms, specifically equivalentClass, equivalentProperty, inverseOf and subClassOf, are employed for the definition of constraints on the score function considered in the training phase so that the resulting vectors, related to such axioms, reflect their specific properties. As a consequence, new triples are also added to the training set on the grounds of the specified axioms.

TRANSROWL further evolves the approach used to derive TRANSOWL from TRANSE by adopting TRANSR as the base model in order to handle non 1-to-1 properties in a more proper way. Indeed, the poor modeling of these relations (caused by TRANSE) may generate spurious embedding vectors with null values or analogous vectors among different entities, thus compromising the ability of making correct predictions. A noteworthy case regards the typeOf property, a common N-to-N relationship. Modeling such property with TRANSE amounts to a simple vector translation; the considered individuals and classes may be quite different in terms of properties and attributes they are involved in, thus determining strong semantic differences (according

[4] TRANSR tackles some weak points in TRANSE, such as the difficulty of modeling specific types of relationships [3].

[5] Facilities available in the Apache Jena framework were used: https://jena.apache.org.

to [76]) taking place at large reciprocal distances in the underlying vector space, hence revealing the weakness of employing the mere translation. Differently, TRANSR associates to typeOf, and to all other properties, a specific vector space where entity vectors are projected to. This leads to training specific projection matrices for typeOf so that the projected entities can be located more suitably to be linked by the vector translation associated to typeOf.

These models, characterized by learning embeddings whilst exploiting prior knowledge both during the learning process and the triple corruption process, have been proved to improve their effectiveness compared to the original models, that focus on structural graph properties with a random corruption process, on link prediction and triple classification tasks. Nevertheless, they also showed some shortcomings since they suffered when some of the considered schema axioms were missing, thus suggesting that further research needs to be pursued in this direction.

3.4 Vector Space Embeddings for Propositionalization

A complementary research direction focused 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 [43], 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 [61] has been proposed. It formalizes a solution for learning latent numeric 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); 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 independent from task and algorithm, an experimental evaluation involving a number of classification and regression tasks has been performed.

An upgrade of RDF2Vec has been presented in [14]. 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 On the Need for Explainable Solutions

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.

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 almost forgetting one of the most powerful characteristic of these languages, that is being empowered with deductive reasoning capabilities that allow for deriving 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 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 the very large amount of information to provide potentially more accurate results.

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

Some discussions in this direction have been developed by the Neural-Symbolic Learning and Reasoning community [30,34], which seeks to integrate principles from neural networks learning and logical reasoning. The main conclusion has been that neural-symbolic integration appears particularly suitable for 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 outlined [30], such as: how to represent expressive logics within neural networks, how neural networks should reason with variables, or how to extract symbolic representation from trained neural networks. Preliminary results for some of these challenges have been recently registered, encouraging pursuing the research direction. An example is represented by SimplE [41], a scalable tensor-based factorization model that is able to learn interpretable embeddings incorporating logical rules through weight tying. Ideas for extracting propositional rules from trained neural networks under SW background knowledge have been illustrated [44], showing that the exploitation of BK allows for: reducing the extracted rule set; reproducing the input-output function of the trained neural network. A conceptual sketch for explaining the classification behavior of artificial neural networks in a non-propositional setting while using SW background knowledge has been proposed [65]. This sheds the light on another important issue, that is the necessity to provide explanations for results supplied by ML methods [15], particularly when they come from very large sources of knowledge, e.g. results for a link prediction problem.

The solution depicted in [65] is in agreement with the idea of exploiting symbol-based interpretable models to explain conclusions [49,58]. Nevertheless, interpretable models describe *how* solutions are obtained but not *why* they are obtained. As argued in [22,30], 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 [13]. It takes advantage of ontologies (DBPedia is used) and reasoning capabilities to infer different kinds of human understandable explanatory evidence. Hence, in a more broad sense, providing an explanation means to open the box of the reasoning process and make it understandable. In a complex setting such as the Web of Data, where knowledge may result from 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. deduction (when rules and theory are available), induction (for building models from the available knowledge), abduction (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.

5 Conclusions

This paper surveyed the different SW problems where ML solutions have been employed and the progresses that have been registered from them. A major focus has been devoted to the main issues that need to be considered and solved when ML solutions are adopted in the SW field. Specifically, symbol-based and numeric-based methods have been analyzed and their main peculiarities and drawbacks have been highlighted. Hence, some considerations concerning the need for solutions that are able to provide human understandable explanations and, towards this direction, to come up with a unified framework integrating both numeric and symbol-based solutions, have been reported.

References

1. Abboud, R., Ceylan, İ.İ., Lukasiewicz, T., Salvatori, T.: BoxE: a box embedding model for knowledge base completion. In: Proceedings of NeurIPS 2020 (2020)
2. Agrawal, R., Imieliński, T., Swami, A.: Mining association rules between sets of items in large databases. In: Buneman, P., et al. (eds.) Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, pp. 207–216. ACM Press (1993). https://doi.org/10.1145/170035.170072
3. Arnaout, H., Razniewski, S., Weikum, G.: Enriching knowledge bases with interesting negative statements. In: Das, D., et al. (eds.) Proceedings of AKBC 2020 (2020). https://doi.org/10.24432/C5101K
4. Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F. (eds.): Description Logic Handbook, 2nd edn. Cambridge University Press, Cambridge (2010). https://doi.org/10.1017/CBO9780511711787
5. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. Sci. Am. **284**(5), 34–43 (2001). https://doi.org/10.4018/jswis.2009081901
6. Bizer, C., Heath, T., Berners-Lee, T.: Linked data - the story so far. Int. J. Sem. Web Inf. Syst. **5**(3), 1–22 (2009). https://doi.org/10.4018/jswis.2009081901
7. Blockeel, H., De Raedt, L.: Top-down induction of first-order logical decision trees. Artif. Intell. **101**(1–2), 285–297 (1998). https://doi.org/10.1016/S0004-3702(98)00034-4

8. Bloehdorn, S., Sure, Y.: Kernel methods for mining instance data in ontologies. In: Aberer, K., et al. (eds.) ASWC/ISWC -2007. LNCS, vol. 4825, pp. 58–71. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-76298-0_5
9. Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., Yakhnenko, O.: Translating embeddings for modeling multi-relational data. In: Burges, C.J.C., et al. (eds.) Proceedings of NIPS 2013, pp. 2787–2795. Curran Associates, Inc. (2013)
10. Cai, H., Zheng, V.W., Chang, K.: A comprehensive survey of graph embedding: Problems, techniques, and applications. IEEE Trans. Knowl. Data Eng. **30**(09), 1616–1637 (2018). https://doi.org/10.1109/TKDE.2018.2807452
11. Carbonneau, M., Cheplygina, V., Granger, E., Gagnon, G.: Multiple instance learning. Pattern Recogn. **77**, 329–353 (2018). https://doi.org/10.1016/j.patcog.2017.10.009
12. Chapelle, O., Schölkopf, B., Zien, A. (eds.): Semi-supervised Learning. The MIT Press, Cambridge (2006). https://doi.org/10.7551/mitpress/9780262033589.001.0001
13. Chen, J., Lécué, F., Pan, J., Horrocks, I., Chen, H.: Knowledge-based transfer learning explanation. In: Thielscher, M., et al. (eds.) Principles of Knowledge Representation and Reasoning: Proceedings of the Sixteenth International Conference, KR 2018, pp. 349–358. AAAI Press (2018)
14. Cochez, M., Ristoski, P., Ponzetto, S.P., Paulheim, H.: Global RDF vector space embeddings. In: d'Amato, C., et al. (eds.) ISWC 2017. LNCS, vol. 10587, pp. 190–207. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-68288-4_12
15. 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), Dagstuhl Reports, vol. 8, pp. 37–38. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik (2019). https://doi.org/10.4230/DagRep.8.9.29
16. d'Amato, C., Fanizzi, N., Esposito, F.: Query answering and ontology population: an inductive approach. In: Bechhofer, S., Hauswirth, M., Hoffmann, J., Koubarakis, M. (eds.) ESWC 2008. LNCS, vol. 5021, pp. 288–302. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-68234-9_23
17. d'Amato, C., Fanizzi, N., Esposito, F.: Inductive learning for the semantic web: what does it buy? Semant. Web **1**(1–2), 53–59 (2010). https://doi.org/10.3233/SW-2010-0007
18. d'Amato, C., Quatraro, N.F., Fanizzi, N.: Injecting background knowledge into embedding models for predictive tasks on knowledge graphs. In: Verborgh, R., et al. (eds.) ESWC 2021. LNCS, vol. 12731, pp. 441–457. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-77385-4_26
19. d'Amato, C., Tettamanzi, A.G.B., Minh, T.D.: Evolutionary discovery of multi-relational association rules from ontological knowledge bases. In: Blomqvist, E., Ciancarini, P., Poggi, F., Vitali, F. (eds.) EKAW 2016. LNCS (LNAI), vol. 10024, pp. 113–128. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-49004-5_8
20. d'Amato, C.: Machine learning for the semantic web: lessons learnt and next research directions. Semant. Web **11**(1), 195–203 (2020). https://doi.org/10.3233/SW-200388
21. Deng, L., Yu, D. (eds.): Deep Learning: Methods and Applications. NOW Publishers, Delft (2014). https://doi.org/10.1561/2000000039
22. Doran, D., Schulz, S., Besold, T.: What does explainable AI really mean? A new conceptualization of perspectives. In: Besold, T.R., Kutz, O. (eds.) Proceedings of the First International Workshop on Comprehensibility and Explanation in AI and ML 2017 co-located with 16th International Conference of the Italian Association for Artificial Intelligence (AI*IA 2017), CEUR Work. Proc., vol. 2071. CEUR-WS.org (2017)
23. Fanizzi, N., d'Amato, C., Esposito, F.: Conceptual clustering and its application to concept drift and novelty detection. In: Bechhofer, S., Hauswirth, M., Hoffmann, J., Koubarakis, M.

(eds.) ESWC 2008. LNCS, vol. 5021, pp. 318–332. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-68234-9_25

24. Fanizzi, N., d'Amato, C., Esposito, F.: DL-FOIL concept learning in description logics. In: Železný, F., Lavrač, N. (eds.) ILP 2008. LNCS (LNAI), vol. 5194, pp. 107–121. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-85928-4_12
25. Fanizzi, N., d'Amato, C., Esposito, F.: Metric-based stochastic conceptual clustering for ontologies. Inf. Syst. **34**(8), 792–806 (2009). https://doi.org/10.1016/j.is.2009.03.008
26. Fanizzi, N., d'Amato, C., Esposito, F.: Induction of concepts in web ontologies through terminological decision trees. In: Balcázar, J.L., Bonchi, F., Gionis, A., Sebag, M. (eds.) ECML PKDD 2010. LNCS (LNAI), vol. 6321, pp. 442–457. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15880-3_34
27. Fanizzi, N., Rizzo, G., d'Amato, C.: Boosting DL concept learners. In: Hitzler, P., et al. (eds.) ESWC 2019. LNCS, vol. 11503, pp. 68–83. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21348-0_5
28. 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.) 22nd International World Wide Web Conference, WWW 2013, pp. 413–422. International World Wide Web Conferences Steering Committee/ACM (2013). https://doi.org/10.1145/2488388.2488425
29. Galárraga, L., Teflioudi, C., Hose, K., Suchanek, F.M.: Fast rule mining in ontological knowledge bases with AMIE+. The VLDB J. **24**(6), 707–730 (2015). https://doi.org/10.1007/s00778-015-0394-1
30. d'Avila Garcez, A., et al.: Neural-symbolic learning and reasoning: contributions and challenges. In: 2015 AAAI Spring Symposia. AAAI Press (2015). http://www.aaai.org/ocs/index.php/SSS/SSS15/paper/view/10281
31. Getoor, L., Taskar, B. (eds.): Introduction to Statistical Relational Learning. MIT Press, Cambridge (2007)
32. Guo, H., Herna, V.L.: Learning from imbalanced data sets with boosting and data generation: the databoost-im approach. SIGKDD Explor. **6**(1), 30–39 (2004). https://doi.org/10.1145/1007730.1007736
33. Guo, S., Wang, Q., Wang, L., Wang, B., Guo, L.: Jointly embedding knowledge graphs and logical rules. In: Proceedings of EMNLP 2016, pp. 192–202. ACL (2016). https://doi.org/10.18653/v1/D16-1019
34. Hitzler, P., Bianchi, F., Ebrahimi, M., Sarker, M.K.: Neural-symbolic integration and the semantic web. Semant. Web J. **11**(1), 3–11 (2020). https://doi.org/10.3233/SW-190368
35. Hoekstra, R.: The knowledge reengineering bottleneck. Semant. Web J. **1**, 111–115 (2010). https://doi.org/10.3233/SW-2010-0004
36. Hogan, A., et al.: Knowledge graphs. ACM Comput. Surv. **54**, 1–37 (2021). https://doi.org/10.1145/3447772
37. Horrocks, I., Patel-Schneider, P., Boley, H., Tabet, S., Grosof, B., Dean., M.: SWRL: a semantic web rule language combining owl and RuleML (2004). http://www.aaai.org/ocs/index.php/SSS/SSS15/paper/view/10281
38. Jayathilaka, M., Mu, T., Sattler, U.: Visual-semantic embedding model informed by structured knowledge. In: Rudolph, S., Marreiros, G. (eds.) Proceedings of STAIRS 2020. CEUR, vol. 2655. CEUR-WS.org (2020). http://ceur-ws.org/Vol-2655/paper23.pdf
39. Ji, S., Pan, S., Cambria, E., Marttinen, P., Yu, P.S.: A survey on knowledge graphs: representation, acquisition and applications (2020). arXiv:2002.00388
40. Józefowska, J., Lawrynowicz, A., Lukaszewski, T.: The role of semantics in mining frequent patterns from knowledge bases in description logics with rules. TPLP **10**(3), 251–289 (2010). https://doi.org/10.1017/S1471068410000098

41. Kazemi, S., Poole, D.: Simple embedding for link prediction in knowledge graphs. In: Bengio, S., et al. (eds.) Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, pp. 4289–4300. ACM (2018)
42. Koller, D., Friedman, N. (eds.): Probabilistic Graphical Models: Principles and Techniques. MIT Press, Cambridge (2009)
43. Kramer, S., Lavrač, N., Flach, P.: Propositionalization approaches to relational data mining. In: Džeroski, S., Lavraž, N. (eds.) Relational Data Mining, pp. 262–291. LNCS, Springer (2001). https://doi.org/10.1007/978-3-662-04599-2_11
44. Labaf, M., Hitzler, P., Evans, A.: Propositional rule extraction from neural networks under background knowledge. In: Besold, T.R., et al. (eds.) Proceedings of the Twelfth International Workshop on Neural-Symbolic Learning and Reasoning, NeSy 2017. CEUR Workshop Proceedings, vol. 2003. CEUR-WS.org (2017)
45. Lehmann, J., Auer, S., Bühmann, L., Tramp, S.: Class expression learning for ontology engineering. J. Web Semant. **9**(1), 71–81 (2011). https://doi.org/10.1016/j.websem.2011.01.001
46. Lehmann, J., Bühmann, L.: ORE - a tool for repairing and enriching knowledge bases. In: Patel-Schneider, P.F., et al. (eds.) ISWC 2010. LNCS, vol. 6497, pp. 177–193. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-17749-1_12
47. Lin, Y., Liu, Z., Sun, M., Liu, Y., Zhu, X.: Learning entity and relation embeddings for knowledge graph completion. In: AAAI 2015 Proceedings, pp. 2181–2187. AAAI Press (2015)
48. Liu, X., Wu, J., Zhou, Z.: Exploratory under-sampling for class-imbalance learning. In: Proceedings of the 6th IEEE International Conference on Data Mining (ICDM 2006), pp. 965–969. IEEE Computer Society (2006). https://doi.org/10.1109/ICDM.2006.68
49. Luger, G.F. (ed.): Artificial Intelligence: Structures and Strategies for Complex Problem Solving. Addison Wesley, Boston (2005)
50. Melo, A., Völker, J., Paulheim, H.: Type prediction in noisy RDF knowledge bases using hierarchical multilabel classification with graph and latent features. Int. J. Artif. Intell. Tools **26**(2), 1–32 (2017). https://doi.org/10.1142/S0218213017600119
51. Minervini, P., d'Amato, C., Fanizzi, N.: Efficient energy-based embedding models for link prediction in knowledge graphs. J. Intell. Inf. Syst. **47**(1), 91–109 (2016). https://doi.org/10.1007/s10844-016-0414-7
52. Minervini, P., Tresp, V., d'Amato, C., Fanizzi, N.: Adaptive knowledge propagation in web ontologies. TWEB **12**(1), 2:1-2:28 (2018). https://doi.org/10.1145/3105961
53. Minervini, P., Costabello, L., Muñoz, E., Nováček, V., Vandenbussche, P.-Y.: Regularizing knowledge graph embeddings via equivalence and inversion axioms. In: Ceci, M., Hollmén, J., Todorovski, L., Vens, C., Džeroski, S. (eds.) ECML PKDD 2017. LNCS (LNAI), vol. 10534, pp. 668–683. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-71249-9_40
54. Minervini, P., Demeester, T., Rocktäschel, T., Riedel, S.: Adversarial sets for regularising neural link predictors. In: Elidan, G., et al. (eds.) UAI 2017 Proceedings. AUAI Press (2017)
55. Nickel, M., Murphy, K., Tresp, V., Gabrilovich, E.: A review of relational machine learning for knowledge graphs. Proc. IEEE **104**(1), 11–33 (2016). https://doi.org/10.1109/JPROC.2015.2483592
56. Nickel, M., Tresp, V., Kriegel, H.: A three-way model for collective learning on multi-relational data. In: Getoor, L., Scheffer, T. (eds.) Proceedings of the 28th International Conference on Machine Learning, ICML 2011, pp. 809–816. Omnipress (2011). https://icml.cc/2011/papers/438_icmlpaper.pdf
57. Paulheim, H.: Make embeddings semantic again! In: Proceedings of the ISWC 2018 P&D-Industry-BlueSky Tracks. CEUR Workshop Proceedings (2018)
58. Raedt, L.D. (ed.): Logical and Relational Learning: From ILP to MRDM (Cognitive Technologies). Springer-Verlag, Berlin (2008)

59. Rettinger, A., Lösch, U., Tresp, V., d'Amato, C., Fanizzi, N.: Mining the semantic web - statistical learning for next generation knowledge bases. Data Mining Knowl. Disc. **24**(3), 613–662 (2012). https://doi.org/10.1007/s10618-012-0253-2
60. Rettinger, A., Nickles, M., Tresp, V.: Statistical relational learning with formal ontologies. In: Buntine, W., Grobelnik, M., Mladenić, D., Shawe-Taylor, J. (eds.) ECML PKDD 2009. LNCS (LNAI), vol. 5782, pp. 286–301. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-04174-7_19
61. Ristoski, P., Paulheim, H.: RDF2Vec: RDF graph embeddings for data mining. In: Groth, P., et al. (eds.) ISWC 2016. LNCS, vol. 9981, pp. 498–514. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46523-4_30
62. Rizzo, G., d'Amato, C., Fanizzi, N., Esposito, F.: Terminological cluster trees for disjointness axiom discovery. In: Blomqvist, E., et al. (eds.) ESWC 2017. LNCS, vol. 10249, pp. 184–201. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-58068-5_12
63. Rizzo, G., Fanizzi, N., d'Amato, C., Esposito, F.: Approximate classification with web ontologies through evidential terminological trees and forests. Int. J. Approx. Reason. **92**, 340–362 (2018). https://doi.org/10.1016/j.ijar.2017.10.019
64. Rizzo, G., Fanizzi, N., d'Amato, C., Esposito, F.: A framework for tackling myopia in concept learning on the web of data. In: Faron Zucker, C., Ghidini, C., Napoli, A., Toussaint, Y. (eds.) EKAW 2018. LNCS (LNAI), vol. 11313, pp. 338–354. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-03667-6_22
65. Sarker, M., Xie, N., Doran, D., Raymer, M., Hitzler, P.: Explaining trained neural networks with semantic web technologies: First steps. In: Besold, T.R., et al. (eds.) Proceedings of the Twelfth International Workshop on Neural-Symbolic Learning and Reasoning, NeSy 2017. CEUR Workshop Proceedings, vol. 2003. CEUR-WS.org (2017)
66. Shadbolt, N., Berners-Lee, T., Hall, W.: The semantic web revisited. IEEE Intell. Syst. **21**(3), 96–101 (2006). https://doi.org/10.1109/MIS.2006.62
67. Siorpaes, K., Hepp, M.: OntoGame: towards overcoming the incentive bottleneck in ontology building. In: Meersman, R., Tari, Z., Herrero, P. (eds.) OTM 2007. LNCS, vol. 4806, pp. 1222–1232. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-76890-6_50
68. Spinosa, E., de Leon Ferreira de Carvalho, A.P., Gama, J.: Olindda: A cluster-based approach for detecting novelty and concept drift in data streams. In: Symposium of Applied Computing: Proceedings of the ACM International Conference, SAC 2007. vol. 1, pp. 448–452. ACM (2007)
69. 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 2014. LNCS, vol. 8465, pp. 333–348. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07443-6_23
70. Tran, A.C., Dietrich, J., Guesgen, H.W., Marsland, S.: An approach to parallel class expression learning. In: Bikakis, A., Giurca, A. (eds.) RuleML 2012. LNCS, vol. 7438, pp. 302–316. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-32689-9_25
71. Tran, A., Dietrich, J., Guesgen, H., Marsland, S.: Parallel symmetric class expression learning. J. Mach. Learn. Res. **18**(64), 1–34 (2017)
72. Völker, J., Niepert, M.: Statistical schema induction. In: ESWC 2011. LNCS, vol. 6643, pp. 124–138. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21034-1_9
73. Völker, J., Fleischhacker, D., Stuckenschmidt, H.: Automatic acquisition of class disjointness. J. Web Semant. **35**(P2), 124–139 (2015). https://doi.org/10.1016/j.websem.2015.07.001
74. West, R., Gabrilovich, E., Murphy, K., Sun, S., Gupta, R., Lin, D.: Knowledge base completion via search-based question answering. In: Chung, C., et al. (eds.) 23rd International World Wide Web Conference, WWW 2014, pp. 515–526. ACM (2014). https://doi.org/10.1145/2566486.2568032

75. Widmer, G., Kubat, M.: Learning in the presence of concept drift and hidden contexts. Mach. Learn. **23**(1), 69–101 (1996)
76. Yang, B., Yih, W., He, X., Gao, J., Deng, L.: Embedding entities and relations for learning and inference in knowledge bases. In: Proceedings of ICLR 2015 (2015)
77. Zhou, Z., Zhang, M.: Multi-label learning. In: Sammut, C., Geoffrey, W. (eds.) Encyclopedia of Machine Learning and Data Mining, pp. 875–881. Springer, Berlin (2017). https://doi.org/10.1007/978-1-4899-7687-1_910

Author Queries

Chapter 4

Query Refs.	Details Required	Author's response
AQ1	Per Springer style, both city and country names must be present in affiliations. Accordingly we have inserted the city and country name "Bari" and "Italy" in "Aff 1". Please check and confirm if the inserted city and country names "Bari" and "Italy" are correct. If not please provide us with the correct city and country names.	