From Shallow to Deep Interactions Between Knowledge Representation, Reasoning and Machine Learning

Kay R. Amel[1]

GDR Aspects Formels et Algorithmiques de l'Intelligence Artificielle
CNRS

December 13, 2019

Abstract

This paper proposes a tentative and original survey of meeting points between Knowledge Representation and Reasoning (KRR) and Machine Learning (ML), two areas which have been developing quite separately in the last three decades. Some common concerns are identified and discussed such as the types of used representation, the roles of knowledge and data, the lack or the excess of information, or the need for explanations and causal understanding. Then some methodologies combining reasoning and learning are reviewed (such as inductive logic programming, neuro-symbolic reasoning, formal concept analysis, rule-based representations and ML, uncertainty in ML, or case-based reasoning and analogical reasoning), before discussing examples of synergies between KRR and ML (including topics such as belief functions on regression, EM algorithm versus revision, the semantic description of vector representations, the combination of deep learning with high level inference, knowledge graph completion, declarative frameworks for data mining, or preferences and recommendation). This paper is the first step of a work in progress aiming at a better mutual understanding of research in KRR and ML, and how they could cooperate.

1   Introduction

Reasoning and learning are two basic concerns at the core of Artificial Intelligence (AI). In the last three decades, Knowledge Representation and Reasoning (KRR) on the one hand, and Machine Learning (ML) on the other hand, have been considerably developed and have specialised themselves in a large number of dedicated sub-fields. These technical developments and specialisations, while they were strengthening the respective corpora of methods in KRR and in ML, also contributed to an almost complete separation of the lines of research in these two areas, making many researchers on one side largely ignorant of what is going on, on the other side.

This state of affairs is also somewhat relying on general, overly simplistic, dichotomies that suggest there exists a large gap between KRR and ML: KRR deals with knowledge, ML handles data; KRR privileges symbolic, discrete approaches, while numerical methods dominate ML. Even if such a rough picture points out facts that cannot be fully denied, it is also misleading, as for instance KRR can deal with data as well [Pra16] (e.g., in formal concept analysis) and ML approaches may rely on symbolic knowledge (e.g., in inductive logic programming). Indeed, the frontier between the two fields is actually much blurrier than it appears, as both are involved in approaches such as Bayesian networks, or case-based reasoning and analogical reasoning, and they share important concerns such as uncertainty representation (e.g., probabilistic or possibilistic models, belief functions, imprecise probability-based approaches).

---

[1]Kay R. Amel is the pen name of the working group "Apprentissage et Raisonnement" of the GDR ("Groupement De Recherche") named "Aspects Formels et Algorithmiques de l'Intelligence Artificielle", CNRS, France (https://www.gdria.fr/presentation/). The contributors to this paper include: Zied Bouraoui (CRIL, Lens, Fr, bouraoui@cril.fr), Antoine Cornuéjols (AgroParisTech, Paris,Fr, antoine.cornuejols@agroparistech.fr), Thierry Denœux (Heudiasyc, Compiègne, Fr, thierry.denoeux@utc.fr), Sébastien Destercke (Heudiasyc, Compiègne, Fr, sebastien.destercke@hds.utc.fr), Didier Dubois (IRIT, Toulouse, Fr, dubois@irit.fr), Romain Guillaume (IRIT, Toulouse, Fr, Romain.Guillaume@irit.fr), João Marques-Silva (ANITI, Toulouse, Fr, joao.marques-silva@univ-toulouse.fr), Jérôme Mengin (IRIT, Toulouse, Fr, Jerome.Mengin@irit.fr), Henri Prade (IRIT, Toulouse, Fr, prade@irit.fr), Steven Schockaert (School of Computer Science and Informatics, Cardiff, UK, SchockaertS1@cardiff.ac.uk), Mathieu Serrurier (IRIT, Toulouse, Fr, mathieu.serrurier@gmail.com), Christel Vrain (LIFO, Orléans,Fr, Christel.Vrain@univ-orleans.fr).

These remarks already suggest that KRR and ML may have more in common than one might think at first glance. In that respect, it is also important to remember that the human mind is able to perform both reasoning and learning tasks with many interactions between these two types of activity. In fact, from the very beginning of the AI history, both reasoning and learning tasks have been considered, but not by the same researchers; see, e.g., [DP19]. So, especially if the ultimate goal of AI is to have machines that perform tasks handled by the human mind, it might be natural and useful to increase the cooperation between KRR and ML.

The intended goal pursued in this work is to start and construct an inventory of common concerns in KRR and ML, of methodologies combining reasoning principles and learning, of examples of KRR/ML synergies. Yet, this paper is not an overview of the main issues in KRR crossed with an overview of the main issues in ML, trying to identify when they meet. Doing so would lead to a huge and endless survey since providing a survey of methodologies and tools for KRR alone, or for ML alone would be already a colossal task[2]. In the following we rather try to identify a collection of meeting points between KRR and ML. Since it is a work in progress, we do not expect to reach any form of exhaustiveness, and even some important topics may remain absent from the document at this stage.

The paper is not either an introductory paper to KRR and/or ML. It is rather intended for readers who are quite familiar with either KRR or ML, and who are curious about the other field. It aims in the long range at contributing to a better mutual understanding of the two communities, and maybe to identify some synergies worth of further research combining KRR and ML.

2   Common Concerns

In order to suggest and illustrate differences and also similarities between KRR and ML, let us start with the simple example of a classification or recommendation-like task ,such as, e.g., associating the profile of a candidate (in terms of skills, tastes) with possible activities suitable for him/her in a vocational guidance system. Such a problem may be envisioned in different manners. On the one hand, one may think of it in terms of a rule-based system relying on some expertise (where rules may be pervaded with uncertainty), or on the other hand in terms of machine learning by exploiting a collection of data (here pertaining to past cases in career guidance).

It is worth noticing that beyond the differences of types of representation that are used in both kinds of approach (e.g., conditional tables for uncertainty assessment vs. weights in a neural net), there are some noticeable similarities between (graphical) structures that can be associated with a rule-based reasoning device, handling uncertainty (or an information fusion process) and with a neural net. This remark suggests that, beyond differences in perspective, there is some structural resemblance between the two types of process. This resemblance has been investigated recently in detail in the setting of belief function theory [Den19], but an example may also be found in an older work on a possibilistic (max-min) matrix calculus devoted to explainability (where each matrix represents a rule) [FP89].

Beyond this kind of parallel, it is clear that KRR and ML have common concerns. This section gives an overview of the main ones regarding the representation issues, the complexity, the role of knowledge, the handling of lack of information, or information in excess, uncertainty, and last but not least regarding causality and explanation. Each subsection below tries to follow the same basic structure, by each time providing i) the KRR view, ii) the ML view, and iii) some synthesis and discussion.

2.1   Types of Representation

In KRR, as suggested by the name, the main representations issues concern the representation of pieces of knowledge (rather than data). The large variety of real world information has led to a number of logical formalisms ranging from classical logic (especially proportional and first order) to modal logics (for dealing with e.g., time, deontic, or epistemic notions) and to non classical logics for handling commonsense reasoning.

---

[2]Nevertheless the interested reader is referred to appropriate chapters in [MPP19] or to monographs such as [BL04, Bar10, Hal17, SSBD14, MRT18].

The representation may use different formats, directed or undirected: sets of if-then rules, or sets of logical formulas. A rule "if A then B" is a 3-valued entity (as first noticed in [DF36]), since it induces a partition between its set of examples, its set of counterexamples and the set of items for which the rule is irrelevant (i.e., when A is false). So a rule strongly departs from its apparent logical counterpart in terms of material implication $A \rightarrow B$ (which is indeed non-directed, since equivalent $\neg B \rightarrow \neg A$). This discrepancy can be also observed in the probabilistic setting, since $Prob(B|A) \neq Prob(A \rightarrow B)$ in general. Rules may hold up to (implicit) exceptions (see subsection 2.3).

Knowledge may be pervaded with uncertainty, which can be handled in different settings, in terms of probability, possibility, belief functions, or imprecise probabilities (see subsection 2.3). In all of these cases, a joint distribution can be decomposed in sub-distributions laying bare some form of conditional independence relations, with a graphical counterpart; the prototypical graphical models in each representation are respectively Bayesian networks (probabilistic), possibilistic networks, credal networks (imprecise probabilities [Coz00]) or valuation-based systems (belief functions). Conceptual graphs [Sow84, CM09] offer a graph representation for logic, especially for ontologies/description logics.

The main goal of KRR is to develop sound and (as far as possible complete) inference mechanisms to draw conclusions from generic knowledge and factual data, in a given representation setting [HFMV03, Hal17]. The mathematical tools underlying KRR are those of logic and uncertainty theories, and more generally discrete mathematics. An important issue in KRR is to find good compromises between the expressivity of the representation and the computational tractability for inferring the conclusions of interest from it [LB87]. This concern is especially at work with description logics that are bound to use tractable fragments of first order logic.

The situation in ML is quite different concerning representation issues. ML aims at learning a model of the world from data. There are thus two key representation problems: the representation of data and the representation of models. See, e.g., [CKN20, CV20]. In many approaches the data space is assimilated to a subset of $\mathbb{R}^p$, in which the observations are described by p numerical attributes. This is the simplest case, allowing the use of mathematical results in linear algebra and in continuous optimization. Nevertheless data may also be described by qualitative attributes, as for instance binary attributes, thus requiring different mathematical approaches, based on discrete optimisation and on enumeration coupled with efficient pruning strategies. Quite often, data is described by both types of attributes and only few ML tools, for instance decision trees, are able to handle them. Therefore, changes of representation are needed, as for instance discretization, or the encoding of qualitative attributes into numerical ones, all inducing a bias on the learning process. More complex data, such as relational data, trees, graphs need more powerful representation languages, such as first order logic or some proper representation trick as for instance propositionalization or the definition of appropriate kernels. It is important to notice that the more sophisticated the representation language, the more complex the inference process and a trade-off must be found between the granularity of the representation and the efficiency of the ML tool.

Regarding models, they depend on the ML task: supervised or unsupervised classification, learning to rank, mining frequent patterns, etc. They depend also on the type of approach that one favours: more statistically or more artificial-intelligence oriented. There is thus a distinction between *generative* and *discriminative* models (or decision functions). In the *generative approach*, one tries to learn a probability distribution $\mathbf{p}_{\mathcal{X}}$ over the input space $\mathcal{X}$. If learning a precise enough probability distribution is successful, it becomes possible in principle to generate further examples $\mathbf{x} \in \mathcal{X}$, the distribution of which is indistinguishable from the true underlying distribution. It is sometimes claimed that this capability makes the generative approach "explicative". This is a matter of debate. The *discriminative* approach does not try to learn a model that allows the generation of more examples. It only provides either a means of deciding when in the supervised mode, or a means to express some regularities in the data set in the unsupervised mode. These regularities as well as these decision functions can be expressed in terms of logical rules, graphs, neural networks, etc. While they do not allow to generate new examples, they nonetheless can be much more interpretable than probability distributions.

Very sketchily, one can distinguish between the following types of representations.

- Linear models and their generalisations, such as linear regression or the linear perceptron first proposed by Rosenblatt [Ros58]. Because these models are based on linear weightings of the descriptors of the entries, it looks easy to estimate the importance of each descriptor and thus to offer some under-

standing of the phenomenon at hand. This, however, assumes that the descriptors are uncorrelated and are well chosen.

- Nonlinear models are often necessary in order to account for the intricacies of the world. Neural networks, nowadays involving very numerous layers of non linearity, are presently the favourite tools for representing and learning non linear models.

- Linear models as well as nonlinear ones provide a description of the world or of decision rules through (finite) combinations of descriptors. They are parametric models. Another approach is to approximate the world by learning a non previously fixed number of prototypes and use a nearest-neighbour technique to define decision functions. These systems are capable of handling any number of prototypes as long as the can fit the data appropriately. *Support Vector Machines* (SVM) fall in this category since they adjust the number of support vectors (learning examples) in order to fit the data. Here, explaining a rule may mean providing a list of the most relevant prototypes that the rule uses.

- The above models are generally numerical in essence, and the associated learning mechanisms most often rely on some optimisation process over the space of parameters. Another class of models relies on logical descriptions, e.g., sets of clauses. Decision trees can also be considered as logic-based, since each tree can be transformed into a set of clauses. The learning algorithms use more powerful structures over the space of models than numerical models. In many cases the discrete nature of the search space and the definition of a generality relation between formulas allow the organization of models in a lattice and the design of heuristics to efficiently prune the search space. More generally, these approaches are usually modeled as enumeration problems (e.g., pattern mining) or discrete optimization problems (supervised learning, clustering, . . .). Moreover such models offer more opportunities to influence the learning process using prior knowledge. Finally, they can be easily interpreted. The downside is their increased brittleness when coping with noisy data.

2.2 Computational Complexity

Complexity issues are a major concern in any branch of computer science. In KRR, very expressive representation languages have been studied, but interesting reasoning problems for these languages are often at least at the second level of the polynomial hierarchy for time complexity. There is a trade-off between the expressive power of a language and the complexity of the inference it allows. Reasoning tasks in languages with suitably restricted expressivity are tractable, like for instance languages using Horn clauses or Lightweight description logics such as DL-lite [CDL⁺05] or EL [BBL05].

The study of complexity has motivated a large number of works in many fields of KRR including non-monotonic reasoning, argumentation, belief merging and uncertainty management. In particular when the desirable solution (i.e., gold standard) of the problem (for instance, merging operator, inconsistency-tolerant consequence relation, etc.) has a high computational complexity, then it is common to look for an approximation that has reasonable complexity. For instance, the observation that answering meaningful queries from an inconsistent DL-Lite knowledge base using universal consequence relation is NP-Complete, has led to the introduction of several tractable approximations [BBB⁺16].

The attempt to cope with hardness of inference has also been a driving force in research around some important and expressive languages, including propositional clauses and CSPs, where inference is NP-complete; for instance, powerful methods nowadays enable the solving of SAT problems with up to hundreds of thousands of variables, and millions of clauses in a few minutes (see section 4.8). Some of the most competitive current SAT solvers are described in [AH15, MML14, LCW⁺15]. Two other ways to cope with time complexity are anytime methods, which can be interrupted at any time during the solving process and then return an incomplete, possibly false or sub-optimal solution; and approximate methods. A recent trend in KRR is to study so-called *compilation schemes* [DM02, Mar15]: the idea here is to pre-process some pieces of the available information in order to improve the computational efficiency (especially, the time complexity) of some tasks; this pre-processing leads to a representation in a language where reasoning tasks can be performed in polynomial time (at the cost of a theoretical blow up in worst-case space complexity, which fortunately does not often happen in practice).

Contrastingly, ML algorithms often have a time complexity which is polynomial in the number of variables, the size of the dataset and the size of the model being learnt, especially when the domains are continuous. However, because of the possible huge size of the dataset or of the models, capping the degree of the polynomial remains an important issue. In the case of discrete domains, finding the optimal model,i.e., the one that best fits a given set of examples, can be hard (see [HR76]), but one is often happy with finding a "good enough" model in polynomial time: there is no absolute guarantee that the model that best fits the examples is the target model anyway, since this may depend on the set of examples. In fact, an important aspect of complexity in ML concerns the prediction of the quality of the model that one can learn from a given dataset: in the PAC setting for instance [Val84], one tries to estimate how many examples are needed to guarantee that the model learnt will be, with a high probability, a close approximation to the unknown target model. Intuitively, the more expressive the hypothesis space is, the more difficult it will be to correctly identify the target model, and the more examples will be needed for that [VC71].

2.3   Lack and Excess of Information: Uncertainty

With respect to a considered reasoning or decision task, information may be missing, or, on the contrary, may be in excess, hence in conflict, which possibly generates uncertainty. Uncertainty has always been an important topic in KRR [Hal17]. While in ML uncertainty is almost always considered to be of statistical or probabilistic origin (aleatory uncertainty), other causes for uncertainty exist, such as the sheer lack of knowledge, and the excess of information leading to conflicts (epistemic uncertainty). However, the role of uncertainty handling in KRR and in ML seems to have been very different so far. While it has been an important issue in KRR and has generated a lot of novel contributions beyond classical logic and probability, it has been considered almost only from a purely statistical point of view in ML [Vap13].

The handling of uncertainty in KRR has a long history, as much with the handling of incomplete information in non-monotonic reasoning as with the handling of probabilities in Bayesian nets [Pea88], and in probabilistic logic languages [Rus15, Coz20]. Other settings that focus on uncertainty due to incomplete information are possibility theory, with weighted logic bases (possibilistic logic [DLP94, DPS17]) and graphical representations (possibilistic nets [BDGP02, BT12]). Belief functions also lend themselves to graphical representations (valuation networks [She94], evidential networks [YM08]) and imprecise probability as well (credal nets [Coz05]).

Uncertainty theories distinct from standard probability theory, such as possibility theory or evidence theory are now well-recognised in knowledge representation. They offer complementary views to uncertainty with respect to probability, or as generalisations of it, dedicated to epistemic uncertainty when information is imprecise or partly missing.

In KRR, at a more symbolic level, the inevitability of partial information has motivated the need for exception-tolerant reasoning. For instance, one may provisionally conclude that "Tweety flies" while only knowing that "Tweety is a bird", although the default rule "birds fly" has exceptions, and we may later conclude that "Tweety does not fly", when getting more (factual) information about Tweety. Thus non-monotonic reasoning [BMT11] has been developed for handling situations with incomplete data, where only plausible tentative conclusions can be derived. Generic knowledge may be missing as well. For example, one may not have the appropriate pieces of knowledge for concluding about some set of facts. Then it may call for interpolation between rules [SP13a].

When information is in excess in KRR, it may mean that it is just redundant, but it becomes more likely that some inconsistency appears. Redundancy is not always a burden, and may sometimes be an advantage by making more things explicit in different formats (e.g., when looking for solutions to a set of constraints).

Inconsistency is a natural phenomenon in particular when trying to use information coming from different sources. Reasoning from inconsistent information is not possible in classical logic (without trivialisation). It has been extensively studied in AI [BH98, BDP97, CC16], in order to try and salvage non-trivial conclusions not involved in contradictions. Inconsistency usually appears at the factual level, for instance a logical base with no model. However, a set of rules may be said to be *incoherent* when there exists an input fact that, together with the rules, would create inconsistency [AR90].

Machine Learning can face several types of situations regarding the amount of information available. It must be said at once that induction, that goes from observations to regularities, is subject to the same

kind of conservation law as in Physics. The information extracted is not created, it is just a reformulation, often with loss, of the incoming information.

If the input data is scarce, then prior knowledge, in one form or another, must complete it. The less data is available, the more prior knowledge is needed to focus the search of regularities by the learning system. This is in essence what the statistical theory of learning says [Vap13]. In recent years, lots of methods have been developed to confront the case where data is scarce and the search space for regularities is gigantic, specially when the number of descriptors is large, often in the thousands or more. The idea is to express special constraints in the so-called regularization term in the inductive criterion that the system use to search the hypothesis space. For instance, a constraint is often that the hypothesis should use a very limited set of descriptors [Tib96].

When there is plenty of data, the problem is more one of dealing with potential inconsistencies. However, except in the symbolic machine learning methods, mostly studied in the 1980s, there is no systematic or principled ways of dealing with inconsistent data. Either the data is pre-processed in order to remove these inconsistencies, and this means having the appropriate prior knowledge to do so, or one relies on the hope that the learning method is robust enough to these inconsistencies and can somehow smooth them up. Too much data may also call for trying to identify a subset of representative data (a relevant sample), as sometimes done in case-based reasoning, when removing redundant cases. Regarding the lack of data there is a variety of approaches for the imputation of missing values ranging from the EM algorithm [DLR77] to analogical proportion-based inference [BPR17]. However these methods get rid of incompleteness and do not reason about uncertainty.

Finally, a situation that is increasingly encountered is that of multi-source data. Then, the characteristics of the multiple data sets can vary, both in the format, the certainty, the precision, and so on. Techniques like data fusion, data aggregation or data integration are called for, often resorting again to prior knowledge, using for instance ontologies to enrich the data.

2.4 Causality and Explainability

"What is an explanation", "What has to be explained, and how" are issues that have been discussed for a long time by psychologists and philosophers [Tha78, Bro92]. The interest in AI for explanations is not new either. It appears with the development of rule-based expert systems in the mid-1980's. Then there was a natural need for explanations that are synthetic, informative, and understandable for the user of an expert system [CDL19]. This raises issues such as designing strategic explanations for a diagnosis, for example in order to try to lay bare the plans and methods used in reaching a goal [HCR84], or using "deep" knowledge for improving explanations [Kas87]. Another issue was the ability to provide negative explanations (as well as positive ones) for answering questions of the form "Why did you not conclude X?" [RS87], even in the presence of uncertainty [FP89].

Introductory surveys about explanations in AI may be found in a series of recent papers [HK17a, HMK17, Kle18, HMM+18, GBY+18, GTFA19]. Let us also mention the problem of explaining the results of a multi-attribute preference model that is like a "black box". It has been more recently studied in [Lab11].

As now discussed, explanations are often related to the idea of causality [HP05]. Indeed most of the explanations we produce or we expect involve some causal relationships (e.g., John imposed on himself to go to the party *because* he thought that Mary would be there). In many domains where machines can provide aid for decision making, as in medicine, court decisions, credit approval and so on, decision makers and regulators more and more want to know what is the basis for the decision suggested by the machine, why it should be made, and what alternative decision could have been made, had the situation been slightly different. One of the difficulties of explainability in Machine learning is due to the fact that algorithms focus on correlations between features and output variables rather than on causality. The example of *wolf vs dog* identification [RSG16] perfectly illustrates this problem. When using a deep classifier, the main feature that determines if a picture represents a dog or a wolf is the presence of snow. There obviously exists a correlation between snow and wolves but it is clearly not a causality link. When this unwanted bias is known, this can be corrected by adding constraint or balancing the dataset. However, due to the lack of interpretability of some algorithms, identifying these biases is challenging. On the other hand, the problem of constraining an algorithm to learn causality rather than correlation is still open.

If explainability is quickly becoming a hot topic in ML, while it was such for expert systems about 30

years ago, the search for solutions is still at an early stage. Some tentative distinction between *interpretability* and *explainability* has been suggested.

Interpretability may have a statistical or a KR interpretation (which are of course not mutually exclusive). From a statistical point of view, an interpretable model is a model that comes with mathematical guarantees. They are usually bounds for the approximation errors (linked to the expression power of the hypothesis space) or the generalization error (linked to the robustness of the algorithm with respect to variations of the sample set). These can be also guarantees about the uncertainty around the parameters of the model (represented by confidence intervals for instance). Linear approaches are, in this scope, the most statistically interpretable ML algorithm. Robustness properties of statistical models are also desirable for interpretable models. This is especially the case when considering explanations based on counterfactual examples. Given a binary classifier and an example *e*, the counterfactual of *e* is the closest example to *e* with respect to a metric that is labeled by the classifier with the opposite label of *e* (the counterfactual is not necessarily in the dataset). Consider for instance a model that determines if a credit is allowed or not with respect to the profile of a customer and a client to which the credit is not granted. The counterfactual in this case answers the question of what is the minimal change on his profile that would ensure that the credit is granted. If the model is based on propositional logic rules, the counterfactual will correspond to a minimal change of the considered example representation in Boolean logic. In this case, the counterfactual is an understandable explanation for the prediction. In deep-learning, the counterpart of counterfactuals are adversarial examples [GSS14]. In most situations, an adversarial example corresponds to an imperceptible modification of the considered example. From this point of view, the lack of robustness of deep networks makes explanations based on counterfactuals very difficult to obtain.

A decision function learned by a system is assumed to be interpretable if it is simple, often meaning a linear model with few parameters, or if it is expressed with terms or rules that a domain expert is supposed to understand, like in rule-based systems or in decision trees. One influencing work promotes the use of locally linear models in order to offer some interpretability even to globally non linear models [RSG16].

Explainability can be understood at two levels. Either at the level of the learning algorithm itself that should be easily understandable by ML experts as well as by practitioners or users of the system. Or at the level of the learned model that, for instance, could incorporate causal relationships. One question is: is it possible to extract causal relationships from data alone, without some prior knowledge that suggest those relationships? Judea Pearl [Pea09, PGJ16, PM18] argues that this is not possible, but gives to the ML techniques the role of identifying possible correlations between variables in huge data sets that are impossible to sift through for human experts. A recent work [LPNC+17] suggests that it would be possible to identify the direction of a causal relationship from observational data. However, the necessary interplay between ML, prior knowledge and reasoning is still a matter of debate.

When dealing with high-dimensional structured data (such as image or text), interpretable and explainable approaches (in a statistical or a KR point of view) are known to be less effective than heavy numerical approaches such as bagging (random forest, gradient boosting) or deep learning [LBH15, GBC16]. Deep learning models are neither explainable nor interpretable due to the large number of parameters and their entanglement. There also exist few statistical results for deep learning and the currently known properties are restricted to very specific architectures (see [ES15] for instance).

Some approaches have been proposed for improving the explainability of a deep learning algorithms (or in some cases any black box algorithm). A first solution for explaining the solution is to analyze the sensitivity of the prediction with respect to small variations of the input. For instance, activation maps [ZKL+15] will emphasise the most important pixel of a picture for a given prediction. Although this type of representation is easily readable, there are some cases where this is not enough for explanation. Another solution is to approximate the model (globally or locally) with an explainable one. Even when the approximation error is reasonable, we have no guarantee that the interpretation associated with the surrogate model is related to the way the initial model makes the prediction. In [LPNC+17], authors propose to locally replace the model with a surrogate interpretable model. This allows to reduce the approximation error but it is based on a neighbourhood notion that can be difficult to define in high-dimensional structured spaces. Moreover, using an interpretable/explainable model is not always a guarantee for explainable prediction. Indeed, a linear function with millions of parameters or a set of rules with thousand

of literals may be not readable at all.

3   Some Methodologies Combining Reasoning Principles and Learning

The idea of combining KRR ingredients with ML tools is not new. It has been done in different ways. This section presents a series of examples of methodologies mixing KRR and ML ideas, with no intent to be exhaustive however. Each subsection roughly follows the same structure, stating the goal of the methodology, presenting its main aspects, and identifying the KR and the ML parts.

3.1   Injecting Knowledge in Learning

Induction cannot be made to work without prior knowledge that restrains the space of models to be explored. Two forms of prior knowledge, aka. biases, are distinguished: *representation biases* that limit the expressiveness of the language used to express the possible hypotheses on the world, and *search biases* that control how the hypothesis space is explored by the learning algorithm.

Representation biases can take various forms. They can directly affect the language in which the possible hypotheses can be expressed. For instance, "hypotheses can involve a maximum of two disjuncts". In the same way, but less declaratively, looking for linear models only is a severe representation bias. Often, one does not want to be so strict, and prefers to favour more flexible preference criteria over the space of possible hypotheses. Generally, this is expressed through a regularized optimisation criterion that balances a measure of fit of the model to the data, and a measure of fit of the model to the bias. For instance, the following quality measure over linear hypotheses $h(x) = \beta_0 + \sum_{j=1}^{p} \beta_j x_{i,j}$ for regression favours hypotheses that involve fewer parameters:

$$R(h) = \underbrace{\frac{1}{2}\sum_{i=1}^{m}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{i,j}\right)^2}_{\text{fit to the data}} + \underbrace{\lambda\sum_{j=1}^{p}||\beta_j||_0}_{\text{Favors models with few non zero parameters}}$$

where the L_0 norm $||.||_0$ counts the nonzero parameters β_j.

The *search bias* dictates how the learning algorithm explores the space of hypotheses. For instance, in the case of neural networks, the search starts with a randomly initialized neural network and then proceeds by a gradient descent optimization scheme. In some other learning methods, such as learning with version space, the search uses generalization relations between hypotheses in order to converge towards good hypotheses. In this latter case, it is more easy to incorporate prior knowledge from the experts. Indeed, the exploration of the hypothesis space is akin to a reasoning process, very much like theorem proving.

3.2   Inductive Logic Programming

Inductive Logic Programming (ILP) (see [MD94, De 08, DL01] for general presentations) is a subfield of ML that aims at learning models expressed in (subsets of) First Order Logic. It is an illustration of Symbolic Learning, where the hypotheses space is discrete and structured by a generality relation. The aim is then to find a hypothesis that covers the positive examples (it is then said to be complete) and rejects the negative ones (it is said to be consistent). The structure of the hypothesis space allows to generalize an incomplete hypothesis, so as to cover more positive examples, or to specialize an inconsistent hypothesis in order to exclude negative covered examples. The main reasoning mechanism is induction in the sense of generalization (subsumption).

In ILP, examples and models are represented by clauses. Relying on First Order Logic allows to model complex problems, involving structured objects (for instance to determine whether a molecule is active or not, a system must take into account the fact it is composed of atoms with their own properties and shared relations), or involving objects in relation with each other (a social network or temporal data). Reasoning is a key part of ILP. First, the search for a model is usually performed by exploring a search space structured by a generality relation. A key point is then the definition of a generality relation between clauses. The more natural definition of subsumption should be expressed in terms of logical consequences, which allows

comparing the models of both formula, but since the problem is in general not decidable, the notion of θ-subsumption, as introduced in [Plo70] is usually preferred: a clause C_1 is more general that a clause C_2 if there exists a substitution θ such that $C_1.\theta \subseteq C_2$. In this definition a clause, i.e. a disjunction of literals, is represented by its set of literals. For instance, the rule $par(X,Y), par(Y,Z) \rightarrow grand\_par(X,Z)$ θ-subsumes $par(john, ann), par(ann, peter), par(john, luc) \rightarrow grand\_par(john, peter)$. Indeed, the first one leads to the clause $\neg par(X,Y) \vee \neg par(Y,Z) \vee grand\_par(X,Z)$ and the second one $\neg par(john, ann) \vee \neg par(ann, peter) \vee \neg par(john, luc) \vee grand\_par(john, peter)$. Second, expert knowledge can be expressed using facts (ground atoms) or by rules, or yet reasoning mechanisms to be applied. This can be illustrated by the well-known systems FOIL [Qui96] and Progol [Mug95]

ILP, and more generally Symbolic Learning, has thus some interesting properties. First, the model is expressed in logic and therefore is claimed to be easily understandable by a user (See for instance [MSZ$^+$18] for an interesting study of the comprehensibility or not of programs learned with ILP). Second, expert knowledge can be easily expressed by means of clauses and integrated into the learning algorithm. Although initially developed for the induction of logic programs, it has now shown its interest for learning with structured data.

However, ILP suffers from two drawbacks: the complexity of its algorithms and its inability to deal with uncertain data. Several mechanisms have been introduced to reduce the complexity, as for instance the introduction of syntactic biases, restricting the class of clauses that can be learned. Another interesting idea is propositionnalization, introduced in [LDG91] and then developed for instance in the system RSD [ZL06]. It is a process that transforms a relational problem into a classical attribute-value problem by the introduction of new features capturing relations between objects. Once the transformation performed, any supervised learner can be applied to the problem. The main difficulty is then to define these new features.

This last problem has led to the emergence of Statistical Relational Learning [GT07, DFKM08] that aims at coupling ILP with probabilistic models. Many systems have been developed, extending naive Bayesian classifier [LF02], Bayesian Networks [FBBR05] or Markov Logic Networks [RD06] or developing new probabilistic framework as in Problog [DKT07]. In all these works, inference and learning are tightly connected since learning parameters requires to maximize the likelihood for generative learning (estimation of the probabilities to generate the data, given a set of parameters), or the conditional likelihood in case of discriminative learning (estimation of the probabilities of the labels given the data). Optimizing the parameters thus require at each step to estimate the corresponding probabilities. This has led to intensive research on the complexity of inference.

In the last decade new works have emerged linking deep learning and Inductive Logic Programming. Two directions are investigated. The first one relies on propositionnalization as in [KKK$^+$17]: first, a set of interpretable rules is built through a Path Ranking Algorithm and then the examples are transformed into an attribute-value representation. Two settings are considered: an existential one making the feature true if an instantiation of this path exists in the example, a counting one that counts the number of times the random walk is satisfied. Once this transformation is performed, a multilayered discriminative RBM can be applied. The second direction, illustrated in [SSZ$^+$17], consists in encoding directly facts and ground rules as neurons. Aggregation neurons allow combining rules with the same head and the activation function are approximations of Łukasiewicz fuzzy logic.

3.3 Neuro-Symbolic Reasoning

Several works have proposed to combine learning and reasoning by studying schemes to translate logical representations of knowledge into neural networks. A long-term goal of a series of works on neural-symbolic integration, surveyed for instance by [BdGB$^+$17], is "to provide a coherent, unifying view for logic and connectionism . . . *[in order to]* . . . produce better computational tools for integrated ML and reasoning." Typical works propose translation algorithms from a symbolic to a connectionist representation, enabling the use of computation methods associated with neural networks to perform tasks associated with the symbolic representation.

Early works in this vein are [Pin91, Pin95, BHKS98, HKS99]. Bornscheuer *et al.* [BHKS98] show for example how an instance of the Boolean satisfiability problem can be translated into a feed-forward network that parallelizes GSAT, a local-search algorithm for Boolean satisfiability. They also show that a normal logic program P can be turned into a connectionnist network that can approximate arbitrarily well

the semantics of the program.

The paper [dGZ99] exploits this idea to represent propositional logic programs with recurrent neural networks (RNNs) which can be used to compute the semantics of the program. They show that this program can also be used as background knowledge to learn from examples, using back-propagation. Essentially, the RNN defined to represent a logic program P has all atoms of P in the input layer; one neuron, a kind of "and" gate, for each rule in a single hidden layer; and one neuron for every atom in the output layer, these neurons working like "or" gates. Re-entrant connections from an atom in the output layer to its counterpart in the input layer enable the chaining of rules.

Franca *et al.* [FZdG14] extend these results to first-order programs, using a propositionalization method called Bottom Clause Propositionalization. In [dGLG07, dGL03], methods are proposed to translate formulas that contain modalities into neural networks, enabling the representation of time and knowledge, and the authors in [dGLG03, dGLG06] show that there exists a neural network ensemble that computes a fixed-point semantics of an intuitionistic theory. Pinkas and Cohen [PC19] perform experiments with so-called higher-order sigma-pi units (which compute a sum of products of their inputs) instead of hidden layers, for planning problems on simple block-world problems: the number of units is fixed at the design time, and is a function of the maximum number of blocks and the maximum number of time steps; for example, for every pair (b_1, b_2) of possible blocks and every time step t, there is a unit representing the proposition above(b_1, b_2, t); their results indicate that a learning phase enables the network to approximately learn the constraints with a reasonable amount of iterations, which speeds up the computation of an approximate solution for subsequent instances.

A new direction of research has recently emerged, bringing in particular to Statistical Relational Learning [GT07, DKNP16] the power of tensor networks. For instance, in [YYC17, HL17], they use recursive tensor networks to predict classes and / or binary relations from a given knowledge base. It has also been proposed in, e.g., [DGMR12, SCMN13, SdG16] to depart from the usual semantics of logic based on Boolean truth values in order to use the numerical operators of neural networks. The universe of an interpretation of first-order logic can be a set of vectors of real numbers, and the truth values of predicates can be real numbers in the interval $[0, 1]$; truth values of general formulas can then be defined using usual operators of fuzzy logic. Donadello *et al.* [DSdG17] describe how this approach can be used to learn semantic image interpretation using background knowledge in the form of simple first-order formulas.

3.4 Formal Concept Analysis

Formal Concept Analysis (FCA) [GW98, FKH$^+$20] is another example of a setting that stands in between KRR and ML concerns. It offers a mathematical framework that is based on the duality between a set of objects or items and a set of descriptors. In the basic setting, we start from a formal context which is a relation linking objects and (Boolean) attributes (or properties). Thus a formal context constitutes a simple repository of data. A concept is formalized as a pair composed of a set of attributes and a set of objects representing the intention and the extension of the concept respectively; it has the property that these objects and only them satisfy the set of attributes and this set of attributes refers to these objects and only them. Such a set of attributes is called a closed pattern or a closed itemset. More precisely, two operators, forming a Galois connection, respectively associate their common descriptors to a subset of objects, and the set of objects that satisfy all of them to a subset of descriptors. In an equivalent way, a pair made of a set of objects and a set of attributes is a formal concept if and only if their Cartesian product forms a maximal rectangle for set inclusion in the formal context. The set of concepts forms a complete lattice.

Formal Concept Analysis is deeply rooted in Artificial Intelligence by the formalization of the notion of concept. Recent years have witnessed a renewed interest in FCA with the emergence of Pattern Mining: it has been shown that the set of closed itemsets forms a condensed representation of the set of itemsets, thus reducing the memory space for storing them. Moreover it a possible to define an equivalence relation between itemsets (two itemsets are equivalent if they share the same closure) and from the equivalence classes, it becomes possible to extract all the exact association rules (rules with a confidence equal to 1). See [GD86, PBTL99, BPT$^+$00]

Two extensions are especially worth mentioning. One uses fuzzy contexts, where the links between objects and attributes are a matter of degree [Bel02]. This may be useful for handling numerical attributes [MDNS08]. Another extension allows for structured or logical descriptors using so-called pattern structures

[GK01, FR04, AKP10]. Besides, operators other than the ones defining formal concepts make sense in formal concept analysis, for instance to characterize independent subcontexts [DP12].

3.5 Rule-Based Models

Knowledge representation by if-then rules is a format whose importance was early acknowledged in the history of AI, with the advent of rule-based expert systems. Their modelling has raised the question of the adequacy of classical logic for representing them, especially in case of uncertainty where conditioning is often preferred to material implication. Moreover, the need for rules tolerating exceptions, or expressing gradedness, such as default rules and fuzzy rules has led KRR to develop tools beyond classical logic.

3.5.1 Default rules

Reasoning in a proper way with default rules (i.e., having potential exceptions) was a challenging task for AI during three decades [BMT11]. Then a natural question is: can rules having exceptions extracted from data be processed by a nonmonotonic inference system yielding new default rules? How can we insure that these new rules are still agreeing with the data? The problem is then to extract genuine default rules that hold in a Boolean database. It does not just amount to mining association rules with a sufficiently high confidence level. We have to guarantee that any new default rule that is deducible from the set of extracted default rules is indeed valid with respect to the database. To this end, we need a probabilistic semantics for nonmonotonic inference. It has been shown [BDP99] that default rules of the form "if p then generally q", denoted by $p \rightsquigarrow q$, where $\rightsquigarrow$ obey the postulates of preferential inference [KLM90], have both

1. a possibilistic semantics expressed by the constraint $\Pi(p \wedge q) > \Pi(p \wedge \neg q)$, for any max-decomposable possibility measure Π ($\Pi(p \vee q) = \max(\Pi(p), \Pi(q))$),

2. a probabilistic semantics expressed by the constraint $Prob(p \wedge q) > Prob(p \wedge \neg q)$ for any *big-stepped probability Prob*.

This is a very special kind of probability such that if $p_1 > p_2 > ... > p_{n-1} \geq p_n$ (where p_i is the probability of one of the n possible worlds), the following inequalities hold $\forall i = 1, n-1$, $p_i > \Sigma_{j=i,n} p_j$. Then, one can safely infer a new default $p \rightsquigarrow q$ from a set of defaults $\Delta = \{p_k \rightsquigarrow q_k | k = 1, K\}$ if and only if the constraints modeling Δ entail the constraints modeling $p \rightsquigarrow q$. Thus, extracting defaults amounts to looking for big-stepped probabilities, by clustering lines describing items in Boolean tables, so as to find default rules, see [BDLP03] for details. Then the rules discovered are genuine default rules that can be reused in a nonmonotonic inference system, and can be encoded in possibilistic logic (assuming rational monotony for the inference relation).

It may be also beneficial to rank-order a set of rules expressed in the setting of classical logic in order to handle exceptions in agreement with nonmonotonic reasoning. This what has been proposed in [SP07] where a new formalization of inductive logic programming (ILP) in first-order possibilistic logic allows us to handle exceptions by means of prioritized rules. Indeed, in classical first-order logic, exceptions of the rules can be assigned to more than one class, even if only one is the right one, which is not correct. The possibilistic formalization provides a sound encoding of non-monotonic reasoning that copes with rules with exceptions and prevents an example from being classified in more than one class.

Possibilistic logic [DLP94] is also a basic logic for handling epistemic uncertainty. It has been established that any set of Markov logic formulas [RD06] can be exactly translated into possibilistic logic formulas [KDS15, DPS17], thus providing an interesting bridge between KRR and ML concerns.

3.5.2 Fuzzy rules

The idea of fuzzy if-then rules was first proposed by L. A. Zadeh [Zad73]. They are rules whose conditions and /or conclusions express fuzzy restrictions on the possible values of variables. Reasoning with fuzzy rules is based on a combination / projection mechanism [Zad79] where the fuzzy pieces of information (rules, facts) are conjunctively combined and projected on variables of interest. Special views of fuzzy rules have been used when designing fuzzy rule-based controllers: fuzzy rules may specify the fuzzy graph of a

control law which once applied to an input yields a fuzzy output that is usually defuzzified [MA75]. Or rules may have precise conclusions that are combined on the basis of the degrees of matching between the current situation and the fuzzy condition parts of the rules [TS85]. In both cases, an interpolation mechanism is at work, implicitly or explicitly [Zad92]. Fuzzy rules-based controllers are universal approximators [Cas95]. The functional equivalence between a radial basis function-based neural network and a fuzzy inference system has been established under certain conditions [JS93].

Moreover, fuzzy rules may provide a rule-based interpretation [dAN94, SHH02] for (simple) neural nets, and neural networks can be used for extracting fuzzy rules from the training data [Ped98, GL05]. Regarding neural nets, let us also mention a non-monotonic inference view [BG91, Gär91].

Association rules [AIS93, HH78] describe relations between variables together with confidence and support degrees. See [DHP06] for the proper assessment of confidence and support degrees in the fuzzy case. In the same spirit, learning methods for fuzzy decision trees have been devised in [MB10], in the case of numerical attributes. The use of fuzzy sets to describe associations between data and decision trees may have some interest: extending the types of relations that may be represented, making easier the interpretation of rules in linguistic terms [DPS05], and avoiding unnatural boundaries in the partitioning of the attribute domains.

There are other kinds of fuzzy rules whose primary goal is not to approximate functions nor to quantify associations, but rather to offer representation formats of interest. This is, for instance, the case of gradual rules, which express statements of the form "the more x is A, the more y is B, where A, and B are gradual properties modelled by fuzzy sets [SDPS07, NLP10].

3.5.3 Threshold rules

Another format of interest is the one of multiple threshold rules, i.e., selection rules of the form "if $x_1 \geq \alpha_1$ and $\cdots x_j \geq \alpha_j$ and $\cdots$ then $y \geq \gamma$" (or deletion rules of the form 'if $x_1 \leq \beta_1$ and $\cdots x_j \leq \beta_j$ and $\cdots$ then $y \geq \delta$"), which are useful in monotone classification / regression problems [GIS06, BSS11]. Indeed when dealing with data that are made of a collection of pairs $(x^k, y_k), k = 1, ..., N$, where x^k is a tuple $(x_1^k, ..., x_n^k)$ of feature evaluations of item k, and where y is assumed to increase with the x_i's in the broad sense, it is of interest of describing the data with such rules of various lengths. It has been noticed [GMS04, DPR14] that, once the numerical data are normalized between 0 and 1, rules where all (non trivial) thresholds are equal can be represented by Sugeno integrals (a generalization of weighted min and weighted max, which is a qualitative counterpart of Choquet integrals [GL10]). Moreover, it has been shown recently [BCD$^+$18] that generalized forms of Sugeno integrals are able to describe a global (increasing) function, taking values on a finite linearly ordered scale, under the form of general thresholded rules. Another approach, in the spirit of the version space approach [Mit79], provides a bracketing of an increasing function by means of a pair of Sugeno integrals [PRSR09, PRS09].

3.6 Uncertainty in ML: in the data or in the model

We will focus on two aspects in which cross-fertilisation of ML with KRR could be envisaged: uncertainty in the data and uncertainty in the models/predictions.

3.6.1 Learning under uncertain and coarse data

In general, learning methods assume the data to be complete, typically in the form of examples being precise values (in the unsupervised case) or precise input/output pairs (in the supervised case). There are however various situations where data can be expected to be uncertain, such as when they are provided by human annotators in classification or measured by low-quality sensors, or even missing, such as when sensors have failed or when only a few examples could be labelled. An important remark is that the uncertainty attached to a particular piece of data can hardly be said to be of objective nature (representing frequency) as it has a unique value, and this even if this uncertainty is due to an aleatory process.

While the case of missing (fully imprecise) data is rather well-explored in the statistical [LR19] and learning [CSZ06] literature, the general case of uncertain data, where this uncertainty can be modelled

using different representation tools of the literature, largely remains to be explored. In general, we can distinguish between two strategies:

- The first one intends to extend the precise methods so that they can handle uncertain data, still retrieving a precise model from them. The most notable approaches consist in either extending the likelihood principle to uncertain data (e.g., [Den13] for evidential data, or [CD18] for coarse data), or to provide a precise loss function defined over partial data and then using it to estimate the empirical risk, see for instance [Hül14, CST11, CS12]. Such approaches are sometimes based on specific assumptions, usually hard to check, about the process that makes data uncertain or partial. Some other approaches such as the evidential likelihood approach outlined in [Den13] do not start from such assumptions, and simply propose a generic way to deal with uncertain data. We can also mention transductive methods such as the evidential K-nearest neighbour (K-NN) rule [Den95, DZ01, DKS19], which allows one to handle partial (or "soft") class labels without having to learn a model.

- The second approach, much less explored, intends to make no assumptions at all about the underlying process making the data uncertain, and considers building the set of all possible models consistent with the data. Again, we can find proposals that extend probability-based approaches [DCZ04], as well as loss-based ones [CS16]. The main criticism one could address to such approaches is that they are computationally very challenging. Moreover they do not yield a single predictive model, making the prediction step potentially difficult and ill-defined, but also more robust.

The problem of handling partial and uncertain data is certainly widely recognised in the different fields of artificial intelligence, be it KRR or ML. One remark is that mainstream ML has, so far, almost exclusively focused on providing computationally efficient learning procedures adapted to imprecise data given in the form of sets, as well as the associated assumptions under which such a learning procedure may work [LD14]. While there are proposals around that envisage the handling of more complex form of uncertain data than just sets, such approaches remain marginal, at least for two major reasons:

- More complex uncertainty models require more efforts at the data collection step, and the benefits of such an approach (compared to set-valued data or noisy precise data) do not always justify the additional efforts. However, there are applications in which the modeling of data uncertainty in the belief function framework does improve the performances in classification tasks [COC$^+$12, QDL17]. Another possibility could be that those data are themselves prediction of an uncertain model, then used in further learning procedures (such as in stacking [DŽ04]);

- Using more complex representations may involve a higher computational cost, and the potential gain of using such representations is not always worth the try. However, some specific cases approaches such as the EM algorithm [Den13] or the K-NN rule [DKS19] in the evidential setting, make it possible to handle uncertain data without additional cost.

3.6.2 Uncertainty in the prediction model

Another step of the learning process where uncertainty can play an important role is in the characterisation of the model or its output values. In the following, we will limit ourselves to the supervised setting where we search to learn a (predictive) function $f : \mathcal{X} \to \mathcal{Y}$ linking an input observation $x \in \mathcal{X}$ to an output (prediction) $y \in \mathcal{Y}$. Assessing the confidence one has in a prediction can be important in sensitive applications. This can be done in different ways:

- By directly impacting the model f itself, for instance associating to every instance x not a deterministic prediction $f(x)$, but an uncertainty model over the domain $\mathcal{Y}$. The most common one is of course probabilities, but other solutions such as possibility distributions, belief functions or convex sets of probabilities are possible;

- By allowing the prediction to become imprecise, the main idea behind such a strategy being to have weaker yet more reliable predictions. In the classical setting, this is usually done by an adequate replacement of the loss function [Ha97, GRKC09], yet recent approaches take a different road. For

instance, imprecise probabilistic approaches consider sets of models combined with a skeptic inference (also a typical approach in KR), where a prediction is rejected if it is so for every possible model [CAZ12]. Conformal prediction [SV08] is another approach that can be plugged to any model output to obtain set-valued predictions. Approaches to quantify statistical predictions in the belief function framework are described in [KSD16, XDZD16, DL18].

If such approaches are relatively well characterised for the simpler cases of multi-class classification, extending them to more complex settings such as multi-label or ranking learning problems that involve combinatorial spaces remain largely unexplored, with only a few contributions [CHWW12, AC17]. It is quite possible that classical AI tools such as SAT or CSP solvers could help to deal with such combinatorial spaces.

3.7 Case-Based Reasoning, Analogical Reasoning and Transfer Learning

Case-based reasoning (CBR for short), e.g., [AP94] is a form of reasoning that exploits data (rather than knowledge) under the form of cases, often viewed as pairs ⟨problem, solution⟩. When one seeks for potential solution(s) to a new problem, one looks for previous solutions to similar problems in the repertory of cases, and then adapts them (if necessary) to the new problem.

Case-based reasoning, especially when similarity is a matter of degree, thus appears to be close to k-NN methods and instance-based learning [Den95, HDP02]. The k-NN method is a prototypical example of transduction, i.e., the class of a new piece of data is predicted on the basis of previously observed data, without any attempt at inducing a generic model for the observed data. The term transduction was coined in [GVV98], but the idea dates back to Bertrand Russell [Rus12].

Another example of transduction is analogical proportion-based learning. Analogical proportions are statements of the form "a is to b as c is to d", often denoted by $a : b :: c : d$, which express that "a differs from b as c differs from d and b differs from a as d differs from c". This statement can be encoded into a Boolean logical expression [MP09, PR13] which is true only for the 6 following assignments $(0,0,0,0)$, $(1,1,1,1)$, $(1,0,1,0)$, $(0,1,0,1)$, $(1,1,0,0)$, and $(0,0,1,1)$ for (a,b,c,d). Note that they are also compatible with the arithmetic proportion definition $a - b = c - d$, where $a - b \in \{-1,0,1\}$, which is not a Boolean expression. Boolean Analogical proportions straightforwardly extend to vectors of attributes values such as $\vec{a} = (a_1, ..., a_n)$, by stating $\vec{a} : \vec{b} :: \vec{c} : \vec{d}$ iff $\forall i \in [1,n]$, $a_i : b_i :: c_i : d_i$. The basic analogical inference pattern [SY05], is then

$$\frac{\forall i \in \{1,...,p\}, \quad a_i : b_i :: c_i : d_i \text{ holds}}{\forall j \in \{p+1,...,n\}, \quad a_j : b_j :: c_j : d_j \text{ holds}}$$

Thus analogical reasoning amounts to finding completely informed triples $(\vec{a}, \vec{b}, \vec{c})$ appropriate for inferring the missing value(s) in $\vec{d}$. When there exist several suitable triples, possibly leading to distinct conclusions, one may use a majority vote for concluding. This inference method extends to analogical proportions between numerical values, and the analogical proportion becomes graded [DPR16]. It has been successfully applied, for Boolean, nominal or numerical attributes, to classification [MBD08, BPR17] (then the class $cl(\vec{x})$ (viewed as a nominal attribute) is the unique solution, when it exists, such as $cl(\vec{a}) : cl(\vec{b}) :: cl(\vec{c}) : cl(\vec{x})$ holds), and more recently to case-based reasoning [LNPR18] and to preference learning [FH18, BPP18]. It has been theoretically established that analogical classifiers *always* yield exact prediction for Boolean affine functions (which includes x-or functions), and only for them [CHPR17]. Good results can still be obtained in other cases [CHPR18]. Moreover, analogical inequalities [PR18b] of the form "a is to b at least as much as c is to d" might be useful for describing relations between features in images, as in [LTC17].

The idea of transfer learning, which may be viewed as a kind of analogical reasoning performed at the meta level, is to take advantage of what has been learnt on a source domain in order to improve the learning process in a target domain related to the source domain. When studying a new problem or a new domain, it is natural to try to identify a related, better mastered, problem or domain from which, hopefully, some useful information can be called upon for help. The emerging area of transfer learning is concerned with finding methods to transfer useful knowledge from a known source domain to a less known target domain.

The easiest and most studied problem is encountered in supervised learning. There, it is supposed that a decision function has been learned in the source domain and that a limited amount of training data is

available in the target domain. For instance, suppose we have learnt a decision function that is able to recognize poppy fields in satellite images. Then the question is: could we use this in order to learn to recognize cancerous cells in biopsies rather than to start anew on this problem or when few labeled data is available in the biological domain?

This type of transfer problem has witnessed a spectacular rise of interest in recent years thanks both to the big data area that makes lots of data available in some domains, and to the onset of deep neural networks. In deep neural networks, the first layers of neuron like elements elaborate on the raw input descriptions by selecting relevant descriptors, while the last layers learn a decision function using these descriptors. Nowadays, most transfer learning methods rely on the idea of transferring the first layers when learning a new neural network on the target training data, adjusting only the last layers. The underlying motivation is that the descriptors are useful in both the source and target domains and what is specific is the decision function built upon these descriptors. But it could be defended just as well that the decision function is what is essential in both domains while the ML part should concentrate on learning an appropriate representation. This has been achieved with success in various tasks [CAMO17].

One central question is how to control what should be transferred. A common assumption is that transfer learning should involve a minimal amount of change of the source domain knowledge in order for it to be used in the target domain. Several ways of measuring this "amount of change" have been put forward (see for instance [CFTR16, SQZY17]), but much work remains to be done before a satisfying theory is obtained.

One interesting line of work is related to the study of causality. Judea Pearl uses the term "transportability" instead of transfer learning, but the fundamental issues are the same. Together with colleagues, they have proposed ways of knowing if and what could be transferred from one domain to another [PM18]. The principles rely on descriptions of the domains using causal diagrams. Thanks to the "do-calculus", formal rules can be used in order to identify what can be used from a source domain to help solve questions in the target domain. One foremost assumption is that causality relationships capture deep knowledge about domains and are somewhat preserved between different situations. For instance, proverbs in natural language are a way of encapsulating such deep causality relationships and their attractiveness comes from their usefulness in many domains or situations, when properly translated.

4 Examples of KRR/ML Synergies

In the previous section, we have surveyed various paradigms where KRR and ML aspects are intricately entwined together. In this section, we rather review examples of hybridizations where KRR and ML tools cooperate. In each case, we try to identify the purpose, the way the KRR and ML parts interact, and the expected benefits of this synergy.

4.1 Dempster-Shafer Reasoning and Generalized Logistic Regression Classifiers

The theory of belief functions originates from Dempster's seminal work [Dem67] who proposed, at the end of the 1960's, a method of statistical inference that extends both Fisher's fiducial inference and Bayesian inference. In a landmark book, Shafer [Sha76] developed Dempster's mathematical framework and extended its domain of application by showing that it could be proposed as a general language to express "probability judgements" (or degrees of belief) induced by items of evidence. This new theory rapidly became popular in Artificial Intelligence where it was named "Dempster-Shafer (DS) theory", evidence theory, or the theory of belief functions. DS theory can be considered from different perspectives:

- A belief function can be defined axiomatically as a Choquet monotone capacity of infinite order [Sha76].

- Belief functions are intimately related to the theory of random sets: any random set induces a belief function and, conversely, any belief function can be seen as being induced by a random set [Ngu78].

- Sets are in one-to-one correspondence with so-called "logical" belief functions, and probability measures are special belief functions. A belief function can thus be seen both as a generalised probability measure and as a generalised set; it makes it possible to combine reasoning mechanisms from

probability theory (conditioning, marginalisation), with set-theoretic operations (intersection, union, cylindrical extension, interval computations, etc.)

DS theory thus provides a very general framework allowing us to reason with imprecise and uncertain information. In particular, it makes it possible to represent states of knowledge close to total ignorance and, consequently, to model situations in which the available knowledge is too limited to be properly represented in the probabilistic formalism. Dempster's rule of combination [Sha76] is an important building block of DS theory, in that it provides a general mechanism for combining independent pieces of evidence.

The first applications of DS theory to machine learning date back to the 1990's and concerned classifier combination [XKS92, Rog94], each classifier being considered as a piece of evidence and combined by Dempster's rule (see, e.g., [QMD11] for a refinement of this idea taking into account the dependence between classifier outputs). In [Den95], Denœux combined Shafer's idea of evidence combination with distance-based classification to introduce the evidential K-NN classifier [Den95]. In this method, each neighbour of an instance to be classified is considered as a piece of evidence about the class of that instance and is represented by a belief function. The K belief functions induced by the K nearest neighbour are then combined by Dempster's rule. Extensions of this simple scheme were later introduced in [ZD98, LRD16, DKS19].

The evidential K-NN rule is, thus, the first example of an "evidential classifier". Typically, an evidential classifier breaks down the evidence of each input feature vector into elementary mass functions and combines them by Dempster's rule. The combined mass function can then be used for decision-making. Thanks to the generality and expressiveness of the belief function formalism, evidential classifiers provide more informative outputs than those of conventional classifiers. This expressiveness can be exploited, in particular, for uncertainty quantification, novelty detection and information fusion in decision-aid or fully automatic decision systems [DKS19].

In [DKS19], it is shown that that not only distance-based classifiers such as the evidential K-NN rule, but also a broad class of supervised machine learning algorithms, can be seen as evidential classifiers. This class contains logistic regression and its non linear generalizations, including multilayer feedforward neural networks, generalized additive models, support vector machines and, more generally, all classifiers based on linear combinations of input or higher-order features and their transformation through the logistic or softmax transfer function. Such *generalized logistic regression classifiers* can be seen as combining elementary pieces of evidence supporting each class or its complement using Dempster's rule. The output class probabilities are then normalized plausibilities according to some underlying belief function. This "hidden" belief function provides a more informative description of the classifier output than the class probabilities, and can be used for decision-making. Also, the individual belief functions computed by each of the features provide insight into the internal operation of classifier and can help to interpret its decisions. This finding opens a new perspective for the study and practical application of a wide range of machine learning algorithms.

4.2 Maximum Likelihood Under Coarse Data

When data is missing or just imprecise (one then speaks of *coarse data*), statistical methods need to be adapted. In particular the question is whether one wishes to model the observed phenomenon *along with* the limited precision of observations, or *despite* imprecision. The latter view comes down to complete the data in some way (using imputation methods). A well-known method that does it is the EM algorithm [DLR77]. This technique makes strong assumptions on the measurement process so as to relate the distribution ruling the underlying phenomenon and the one ruling the imprecise outcomes.It possesses variants based on belief functions [Den13]. EM is extensively used for clustering (using Gaussian mixtures) and learning Bayesian nets.

However the obtained result, where by virtue of the algorithm, data has become complete and precise, is not easy to interpret. If we want to be faithful to the data and its imperfections, one way is to build a model that accounts for the imprecision of observations, i.e., a set-valued model. This is the case if a belief function is obtained via maximum likelihood on imprecise observations: one optimises the *visible likelihood function* [CD18]. The idea is to cover all precise models that could have been derived, had the data been precise. Imprecise models are useful to lay bare ignorance when it is present, so as to urge finding more

data, but it may be problematic for decision or prediction problems, when we have to act or select a value despite ignorance.

Ideally we should optimize the likelihood function based on the actual values hidden behind the imprecise observations. But such a likelihood function is ill-known in the case of coarse data [CD18]. In that case, we are bound

- To make assumptions on the measurement process so as to create a tight link between the hidden likelihood function pertaining to the outcomes of the real phenomenon, and the visible likelihood of the imprecise observations (for instance the CAR (coarsening at random) assumption [HR91], or the superset assumption [HC15]. In that case, the coarseness of the data can be in some sense ignored. See [Jae05] for a general discussion.

- Or to pick a suitable hidden likelihood function among the ones compatible with the imprecise data, for instance using an optimistic maximax approach that considers that the true sample is the best possible sample in terms of likelihood compatible with the imprecise observation [Hül14]. This approach chooses a compatible probability distribution with the minimum of entropy, hence tends to disambiguate the data. On the contrary maximin approach considers that the true sample is the worst compatible sample in terms of likelihood. This approach chooses a compatible probability distribution with the maximum of entropy. Those two approaches suppose extreme point of view on the entropy of the probability distribution. More recently, an approach base on the likelihood ratio that maximize the minimal possible ratio over the compatible probability distribution is proposed in [GD18]. This approach achieving a trade-off between these two more extreme approaches and is able to quantify the quality of the chosen probability distribution in regards to all possible probability distribution. In these approaches, the measurement process is ignored.

See [CDH17, HDC19] for more discussions about such methods for statistical inference with poor quality data.

Besides, another line of work for taking into account the scarcity of data in ML is to use a new cumulative entropy-like function that together considers the entropy of the probability distribution and the uncertainty pertaining to the estimation of its parameters. It takes advantage of the ability of a possibility distribution to upper bound a family of probabilities previously estimated from a limited set of examples [SP13b, SP15]. Such a function takes advantage of the ability of a possibility distribution to upper bound a family of probabilities previously estimated from a limited set of examples and of the link between possibilistic specificity order and entropy [DH07]. This approach enables the expansion of decision trees to be limited when the number of examples at the current final nodes is too small.

4.3 EM Algorithm and Revision

Injecting concepts from KRR, when explaining the EM algorithm may help better figure out what it does. In the most usual case, the coarse data are elements of a partition of the domain of values of some hidden variable. Given a class of parametric statistical models, the idea is to iteratively construct a precise model that fits the data as much as possible, by first generating at each step a precise observation sample in agreement with the incomplete data, followed by the computation of a new model obtained by applying the maximum likelihood method to the last precise sample. These two steps are repeated until convergence to a model is achieved.

In [CD16], it has been shown that the observation sample implicitly built at each step can be represented by a probability distribution on the domain of the hidden variable that is in agreement with the observed frequencies of the coarse data. It is obtained by applying, at each step of the procedure, the oldest (probabilistic) revision rule well-known in AI, namely Jeffrey's rule [Jef83], to the current best parametric model. This form of revision considers a prior probability $p(x, \theta)$ on the domain $\mathcal{X}$ of a variable X, and new information made of a probability distribution over a partition $\{A_1, A_2, \ldots, A_n\}$ of $\mathcal{X}$ (representing the coarse data). If p'_i is the "new" probability of A_i, the old distribution $p(x, \theta)$ is revised so as to be in agreement with the new information. The revised probability function is of the form $p'(x, \theta) = \sum_{i=1}^{n} p'_i \cdot p(x, \theta | A_i)$.

The revision step minimally changes the prior probability function in the sense of Kullback-Leibler relative entropy.

In the case of the EM algorithm, p_i' is the frequency of the coarse observation A_i, and $p(x, \theta)$ is the current best parametric model. The distribution $p'(x, \theta)$ corresponds to a new sample of X in agreement with the coarse observation. In other words the EM algorithm in turn revises the parametric model to make it consistent with the coarse data, and applies maximum entropy to the new obtained sample, thus minimizing the relative (entropic) distance between a parametric model and a probability distribution in agreement with the coarse data.

4.4   Conceptual Spaces and the Semantic Description of Vector Representations

Neural networks, and many other approaches in machine learning, crucially rely on vector representations. Compared to symbolic representations, using vectors has many advantages (e.g., their continuous nature often makes optimizing loss functions easier). At the same time, however, vector representations tend to be difficult to interpret, which makes the models that rely on them hard to explain as well. Since this is clearly problematic in many application contexts, there has been an increasing interest in techniques for linking vector spaces to symbolic representations. The main underlying principles go back to the theory of *conceptual spaces* [Gär00], which was proposed by Gärdenfors as an intermediate representation level between vector space representations and symbolic representations. Conceptual spaces are essentially vector space models, as each object from the domain of discourse is represented as a vector, but they differ in two crucial ways. First, the dimensions of a conceptual space usually correspond to interpretable salient features. Second, (natural) properties and concepts are explicitly modelled as (convex) regions. Given a conceptual space representation, we can thus, e.g., enumerate which properties are satisfied by a given object, determine whether two concepts are disjoint or not, or rank objects according to a given (salient) ordinal feature.

Conceptual spaces were proposed as a framework for studying cognitive and linguistic phenomena, such as concept combination, metaphor and vagueness. As such, the problem of learning conceptual spaces from data has not received much attention. Within a broader setting, however, several authors have studied approaches for learning vector space representations that share important characteristics with conceptual spaces. The main focus in this context has been on learning vector space models with interpretable dimensions. For example, it has been proposed that non-negative matrix factorization leads to representations with dimensions that are easier to interpret than those obtained with other matrix factorization methods [LS99], especially when combined with sparseness constraints [Hoy04]. More recently, a large number of neural network models have been proposed with the aim of learning vectors with interpretable dimensions, under the umbrella term of *disentangled representation learning* [CDH+16, HMP+17]. Another possibility, advocated in [DS15], is to learn (non-orthogonal) directions that model interpretable salient features within a vector space whose dimensions themselves may not be interpretable. Beyond interpretable dimensions, some authors have also looked at modelling properties and concepts as regions in a vector space. For example, [Erk09] proposed to learn region representations of word meaning. More recent approaches along these lines include [VM15], where words are modelled as Gaussian distributions, and [JS17], where word regions were learned using an ordinal regression model with a quadratic kernel. Some authors have also looked at inducing region based representations of concepts from the vector representations of known instances of these concepts [BS18, GBP18]. Finally, within a broader setting, some approaches have been developed that link vectors to natural language descriptions, for instance linking word vectors to dictionary definitions [HCKB16] or images to captions [KFF15].

The aforementioned approaches have found various applications. Within the specific context of explainable machine learning, at least two different strategies may be pursued. One possibility is to train a model in the usual way (e.g., a neural network classifier), and then approximate this model based on the semantic description of the vector representations involved. For instance, [AKS16] suggests a method for learning a rule based classifier that describes a feedforward neural network. The second possibility is to extract an interpretable (qualitative) representation from the vector space, e.g., by treating interpretable dimensions as ordinal features, and then train a model on that interpretable representation [DS15].

4.5 Combining Deep Learning with High Level Inference

While neural networks are traditionally learned in a purely data-driven way, several authors have explored approaches that are capable of taking advantage of symbolic knowledge. One common approach consists in relaxing symbolic rules, expressing available background knowledge, using fuzzy logic connectives. This results in a continuous representation of the background knowledge, which can then simply be added to the loss function of the neural network model [DRR16]. Rather than directly regularizing the loss function in this way, [HML+16] proposes an iterative method to ensure that the proportion of ground instances of the given rules that are predicted to be true by the neural network is in accordance with the confidence we have in these rules. To this end, after each iteration, they solve an optimisation problem to find the set of predictions that is closest to the predictions of the current neural network while being in accordance with the rules. The neural network is subsequently trained to mimic these regularized predictions. Yes another approach is proposed in [XZF+18], which proposes a loss function that encourages the output of a neural network to satisfy a predefined set of symbolic constraints, taking advantage of efficient weighted model counting techniques.

While the aforementioned approaches are aimed at using rules to improve neural network models, some authors have also proposed ways in which neural networks can be incorporated into symbolic formalisms. One notable example along these lines is the DeepProbLog framework from [MDK+18a], where neural networks are essentially used to define probabilistic facts of a probabilistic logic program. Within a broader setting, vector space embeddings have also been used for predicting plausible missing rules in ontological rule bases [BS19, LBS19].

4.6 Knowledge Graph Completion

Knowledge graphs are a popular formalism for expressing factual relational knowledge using triples of the form (entity, relation, entity). In application fields such as natural language processing, they are among the most widely used knowledge representation frameworks. Such knowledge graphs are almost inevitably incomplete, however, given the sheer amount of knowledge about the world that we would like to have access to and given the fact that much of this knowledge needs to be constantly updated. This has given rise to a wide range of methods for automatic knowledge graph completion. On the one hand, several authors have proposed approaches for automatically extracting missing knowledge graph triples from text [RYM10]. On the other hand, a large number of approaches have been studied that aim to predict plausible triples based on statistical regularities in the given knowledge graph. Most of these approaches rely on vector space embeddings of the knowledge graph [BUGD+13, TWR+16]. The main underlying idea is to learn a vector **e** of typically a few hundred dimensions for each entity e, and a scoring function s_R for each relation R, such that the triple (e, R, f) holds if and only if $s_R(\mathbf{e}, \mathbf{f}) \in \mathbb{R}$ is sufficiently high. Provided that the number of dimensions is sufficiently high, any knowledge graph can in principle be modelled exactly in this way [KP18]. To a more limited extent, such vector representations can even capture ontological rules [GS18]. In practice, however, our aim is usually not to learn an exact representation of the knowledge graph, but to learn a vector representation which is predictive of triples that are plausible, despite not being among those in the given knowledge graph. Some authors have also proposed methods for incorporating textual information into knowledge graph embedding approaches. Such methods aim to learn vector space representations of the knowledge graph that depend on both the given knowledge graph triples and textual descriptions of the entities [ZZW+15, JS16, XLJ+16, XHMZ17], or their relationships [RYMM13, TCP+15].

4.7 Declarative Frameworks for Data Mining and Clustering

Machine Learning and Data Mining can also be studied from the viewpoint of problem solving. From this point of view, two families of problems can be distinguished: enumeration and optimisation, the latter being either discrete or continuous.

Pattern mining is the best known example of enumeration problems, with the search for patterns satisfying some properties, as for instance to be frequent, closed, emergent . . . Besides, supervised classification is seen as the search for a model minimizing a given loss function, coupled to a regularization term for avoiding over-fitting, whereas unsupervised learning is modeled as the search of a set of clusters (a par-

tition in many cases) optimizing a quality criterion (the sum of squared errors for instance for k-means). For sake of complexity optimisation problems usually rely on heuristic, with the risk of finding only a local optimum. All these approaches suffer from drawbacks. For instance in pattern mining the expert is often drowned under all the patterns satisfying the given criteria. In optimisation problems, a local optimum can be far from the expert expectations.

To prevent this, the notion of Declarative Data Mining has emerged, allowing the experts to express knowledge in terms of constraints on the desired models. It can be seen as a generalization of semi-supervised classification, where some points are already labelled with classes. Classical algorithms must then be adapted to take into account constraints and that has led to numerous extensions. Nevertheless, most extensions are dedicated to only one type of constraints, since the loss function has to be adapted to integrate their violation and the optimization method (usually a gradient descent) has to be adapted to the new optimization criterion. It has been shown in [DGN08] that declarative frameworks, namely Constraint Programming in that paper, allow to model and handle different kinds of constraints in a generic framework, with no needs to rewrite the solving algorithm. This has been applied to pattern mining and then extended to k-pattern set mining with different applications, such as conceptual clustering or tiling [DGN10, KBC10].

This pioneering work has opened the way to a new branch of research, mainly in Pattern mining and in Constrained Clustering. In this last domain, the constraints were mainly pairwise, e.g., a Mustlink (resp. Cannotlink) constraint expresses that two points must (resp. must not) be in the same cluster. Other constraints have been considered such as cardinality constraints on the size of the clusters, minimum split constraints between clusters. Different declarative frameworks have been used, as for instance SAT [DRS10, JSS17], CP [DDV17], Integer Linear Programming [MK10, BGN14, KRD+17, OZL+17]. An important point is that such frameworks allow to easily embed symbolic and numerical information, for instance by considering a continuous optimisation criterion linked with symbolic constraints, or by considering two optimisation criteria and building a Pareto front [KRD+17].

Thus, declarative frameworks not only allow to easily integrate constraints in Machine Learning problems, but they enable the integration of more complex domain knowledge that goes beyond classical Machine Learning constraints, thus integrating truly meaningful conntraints [DVDD16]. Moreover, new use case for clustering can be considered as for instance given a clustering provided by an algorithm, find a new clustering satisfying new expert knowledge, modifying a minima the previous clustering. The price to pay is computational complexity, and the inability to address large datasets. A new research direction aims at studying how constraints could be integrated to deep learner [ZBD19]. Besides, the Constraint Programming community has also benefited from this new research direction by the development of global constraints tailored to optimize the modeling of Data Mining tasks, as for instance [KLL+17].

4.8   Machine Learning vs. Automated Reasoners

This subsection provides a brief overview of the applications of ML in practical Automated Reasoners, but also overviews the recent uses of Automating Reasoning in ML[1]. With a few exceptions, the subsection emphasizes recent work, published over the last decade. Tightly related work, e.g. inductive logic programming (see Section 3.2) or statistical relational learning [DKNP16], is beyond the scope of this subsection. The subsection is organized in three main parts. First, subsubsection 4.8.1 overviews the applications of ML in developing and organizing automated reasoners. Second, subsubsection 4.8.2 covers the recent users of automated reasoning in learning ML models, improving the robustness of ML models, but also in explaining ML models. Finally, subsubsection 4.8.3 covers a number of recent topics at the intersection of automated reasoning and ML.

4.8.1   Learning for Reasoning

Until recently, the most common connection between ML and automated reasoning would be to apply the former when devising solutions for the latter. As a result, a wealth of attempts have been made towards

---

[1]We adopt a common understanding of *Automated Reasoning* as "*The study of automated reasoning helps produce computer programs that allow computers to reason completely, or nearly completely, automatically*" (from https://en.wikipedia.org/wiki/Automated_reasoning).

applying ML in the design of automated reasoners, either for improving existing algorithms or for devising new algorithms, built on top of ML models. Uses of ML can be organized as follows. First, uses of ML for improving specific components of automated reasoners, or for automatic configuration or tuning of automated reasoners. Second, approaches that exploit ML for solving computationally hard decision, search and counting problems, and so offering alternatives to dedicated automated reasoners.

Improving Reasoners.   Earlier efforts on exploiting ML in automated reasoners was to improve specific components of reasoners by seeking guidance from some ML model. A wealth of examples exist, including the improvement of restarts in Boolean Satisfiability (SAT) solvers [HW09], improvement of branching heuristics [FB12, GP14, LGPC16b, LGPC16a, LOM$^+$18], selection of abstractions for Quantified Boolean Formulas (QBF) solving [Jan18, LRS18], but also for improving different components of theorem provers for first-order and higher order logics [UVS11, KU14, KUV14, KU15, ISA$^+$16, LISK17, CJSU19].

ML has found other uses for improving automated reasoners. A well-known example is the organization of portfolio solvers [XHHL08, HHL11, HXHL14, KXHL16, MPLM08]. Another example is the automatic configuration of solvers, when the number of options available is large [HHLS09, SNGR09]. One additional example is the automatic building of automated reasoners using ML [KXHL16].

Tackling Computationally Hard Problems.   Another line of work has been to develop solutions for solving computationally hard decision and search problems. Recent work showed promise in the use of NNs for solving satisfiable instances of SAT represented in clausal form [SLB$^+$18, SLB$^+$19, SB19, WDWK19], for solving instances of SAT represented as circuits [AMW19a, AMW19b], but also NP-complete problems in general [PAL$^+$19]. The most often used approach has been to exploit variants of Graph Neural Networks (GCNs) [SGT$^+$09], including Message Passing Neural Networks (MPNNs) [GSR$^+$17]. There has also been recent work on solving CSPs [XKK18] using convolutional NNs. Furthermore, there have been proposals for learning to solve SMT [BBV18], combinatorial optimization problems [KDZ$^+$17, BPL$^+$17, LCK18, BLP18], planning problems [DBW18], but also well-known specific cases of NP-complete decision problems, e.g. Sudoku [PPW18] and TSP [VFJ15].

Efforts for tackling computationally harder problems have also been reported, including QBF [YWC$^+$19], ontological reasoning [HL17], probabilistic logic programming [MDK$^+$18b], inference in probabilistic graphical models [YLX$^+$18] and theorem proving for first order [KUMO18, HDSS19, BLR$^+$19, PLR$^+$19, Hua19] and higher order logics [Wha16, WTWD17, KCS17, YD19].

4.8.2   Reasoning for Learning

This subsection overviews uses of automated reasoning approaches for verifying, explaining and learning ML models.

Robust Machine Learning.   Concerns about the behavior of neural networks can be traced at least to the mid 90s and early 00s [WCN96, Zak01, SN02]. Additional early work on ensuring safety of neural networks also involved SAT solvers [PT10]. More recently, efforts on the verification of neural networks have focused on the avoidance of so-called adversarial examples.

Adversarial examples [SZS$^+$14], already briefly mentioned in Subsection 2.4, illustrate the brittleness of ML models. In recent years, a number of unsettling examples served to raise concerns on the fragility neural networks can be in practice [AFGG17, EEF$^+$18, CAD$^+$18, FBI$^+$19, Hea19]. Among other alternative approaches, automated reasoners have been applied to ensuring the robustness of ML models, emphasizing neural networks. A well-known line of work focuses on the avoidance of adversarial examples for neural networks using ReLU units [NH10] and proposes Reluplex, an SMT-specific dedicated reasoning engine for implementing reasoning with ReLU units [KBD$^+$17, GKPB18, KHI$^+$19]. Another recent line of work addresses binarized neural networks [HCS$^+$16] and develops a propositional encoding for assessing the robustness of BNNs [NKR$^+$18, Nar18, LNPT18]. Additional lines of work have been reported in the last two years [HKWW17, RHK18, CKL$^+$19, Kwi19, GMD$^+$18, MGV18, SGM$^+$18, SGPV19b, SGPV19a, BMV19, DSG$^+$18, BTT$^+$18, DGFK18, QDO$^+$19, DSG$^+$19, WBD$^+$19, BTT$^+$17, WRTK19, BLT$^+$19, SDD$^+$18, DFG$^+$19, BSS$^+$19].

Interpretable ML Models.   Interpretable ML models are those from which rule-like explanations can be easily produced. For example, decision trees, decision sets (or rule sets) and rule lists are in general deemed interpretable, since one can explain predictions using rules. On area of research is the learning (or synthesis) of interpretable ML models using automated reasoning approaches. There has been continued efforts at learning decision trees [NF07, Nij08, BHO09, NF10, VZ17, NIPM18, VZ19, VNP⁺19, HRS19], decision sets [LBL16, IPNM18, MM18, GM19] and rule lists [ALA⁺17a, ALA⁺17b]. Examples of reasoners used include SAT, CP, and ILP solvers, but dedicated complete methods based on branch and bound search have also been considered. Despite a recent explosion of works on black-box ML models, there exist arguments for the use of interpretable models [Rud19].

Explanations with Abductive Reasoning.   In many settings, interpretable models are not often the option of choice, being replaced by so-called black-box models, which include any ML model from which rules explaining predictions are not readily available. [2]   Concrete examples include (deep) neural networks (including binarized versions), and boosted trees and random forests, among many other alternatives.

Most existing works on computing explanations resort to so-called *local* explanations. These models are agnostic and heuristic in nature [RSG16, LL17, RSG18]. Recent works [NSM⁺19, INM19b] revealed that local explanations do not hold globally, i.e., it is often the case that there are points in feature space, for which the local explanation holds, but for which the model's prediction differs. Since 2018, a number of attempts have been reported, which propose rigorous approaches for computing explanations. Concretely, these recent attempts compute so-called *abductive* explanations, where each explanation corresponds to a prime implicant of the discrete function $\mathcal{F}$ representing the constraint that the ML model predicts the target prediction. A first attempt based on compiling such a function into a tractable representation is reported elsewhere [SCD18]. For such a representation, (shortest) prime implicants can then be extracted in polynomial time. The downside of this approach is that compilation may yield exponential size function representations. Another attempt [INM19a] is based on computing explanations of demand, by encoding the instance, the ML model and the prediction into some logic representation. In this case, reasoners such as SMT, ILP or SAT solvers are then used for extracting (shortest) prime implicants.

Explanations vs. Adversarial Examples.   In recent years, different works realized the existence of some connection between adversarial examples (AE's) and explanations (XP's) [LYH18, TMLZ18, RD18, TWX⁺18, XLZ⁺18, PR18a, CJSW18]. Nevertheless, a theoretical connection between AE's and XP's has been elusive. Recent work [INM19a] showed that adversarial examples can be computed from the set of explanations for some prediction. Furthermore, this work introduced the concept of counterexample (CEx) to some prediction, and identified a minimal hitting set relationship between XP's and CEx's, i.e., XP's are minimal hitting sets of CEx's and vice-versa.

4.8.3   More on Learning vs. Reasoning

The previous two subsections summarize recent efforts on using machine learning for automated reasoning, but also on using automated reasoning for learning, verifying and explaining ML models. This subsection identifies additional lines of research at the intersection of ML and automated reasoning.

Integrating Logic Reasoning in Learning.   A large body of work has been concerned with the integration of logic reasoning with ML. One well-known example is neural-symbolic learning [GBG02, dGLG09, BdGB⁺17, PMS⁺17, BKS18, MGK⁺19, dGGL⁺19]. See also Subsection 3.3. Examples of applications include program synthesis [ZRF⁺18b, ZRF⁺18a, PMS⁺17, BHD⁺18] and neural theorem proving [MBRR18]. Other approaches do exist, including deep reasoning networks [CBZ⁺19], neural logic machines [DML⁺19], and abductive learning [DXYZ18, Zho19]. An alternative is to embed symbolic knowledge in neural networks [XXM⁺19].

---

[2]The definition of *explanation* is the subject of ongoing debate [Mil19b]. We use the intuitive notion of explanation as a IF-THEN rule [RSG16, LL17, RSG18], where some given prediction is made if a number of feature values holds true. The importance of reasoning about explanations is illustrated by a growing number of recent surveys [HK17b, HMK17, BC17, MSM18, Kle18, HMM⁺18, AB18, ACM18, DBH18, HMKL18, GMR⁺19, SMV⁺19, SM19, Mil19b, Mil19a, ANCF19, MRW19, XUD⁺19, MHC⁺19].

Learning for Inference. One area of work is the use of ML models for learning logic representations, most often rules [RR16, RR17, YYC17, EG18b, EG18a], which can serve for inference or for explaining predictions.

Understanding Logic Reasoning. A natural question is whether ML systems understand logical formulas in order to decide entailment or unsatisfiability. There has been recent work on understanding entailment [ESA⁺18, SGHK19], suggesting that this is not always the case, e.g., for convolutional NNs. In a similar fashion, recent work [CY19] suggests that GNNs mail fail at deciding unsatisfiability.

Synthesis of ML Models. Recent work proposed the use of automated reasoners for the synthesis (i.e., learning) of ML models. Concrete examples include [FBD⁺19, ZXMJ19, LZSJ19]. These approaches differ substantially from approaches for the synthesis of interpretable models, including decision trees and sets and decision lists.

As witnessed by the large bibliography surveyed in this subsection, the quantity, the breadth and the depth of existing work at the intersection between ML and automated reasoning in recent years, provides ample evidence that this body of work is expected to continue to expand at a fast pace in the near future.

4.9 Preferences and Recommendation

In both KRR and ML, models have evolved from binary ones (classical propositional or first-order logic, binary classification) to richer ones that take into account the need to propose less drastic outputs. One approach has been to add the possibility to *order* possible outputs / decisions. In multi-class classification tasks for instance, one approach [DKC⁺03] is to estimate, given an instance, the *posterior* probability of belonging to each possible class, and predict the class with highest probability. The possibility of learning to "order things" has numerous applications, e.g., in information retrieval, recommender systems. In KRR, the need to be able to order interpretations (rather than just classify them as possible / impossible, given the knowledge at hand) has proved to be an essential modelling paradigm, see, e.g., the success of valued CSPs [SFV95], Bayesian networks [Pea88], possibilistic / fuzzy logics among others.

At the intersection of ML and KRR, the field of "preference learning" has emerged. Furnkranz et al. [FH10, FHR⁺14] describe various tasks that can be seen as preference learning, some where the output is a function that orders possible labels for each unseen instance, and some where the output is a function that orders any unseen set of new instances.

The importance of the notion of preference seems to have emerged first in economics and decision theory, and research in these fields focused essentially on *utilitarian* models of preferences, where utility function associates a real number with each one of the objects to be ordered. Also in this field research developed on *preference elicitation*, where some interaction is devised to help a decision maker form / lay bare her preferences, usually over a relatively small set of alternatives, possibly considering multiple objectives.

In contrast, preferences in AI often bear on combinatorial objects, like models of some logical theory to indicate for instance preferences over several goals of an agent; or, more recently, like combinations of interdependent decisions or configurable items of some catalog. Thus, in KRR as well as in ML, the objects to be ordered are generally characterised by a finite number of features, with a domain / set of possible values for each feature. When talking about preferences, the domains tend to be finite ; continuous domains can be discretised.

Because of the combinatorial nature of the space of objects, research in AI emphasized the need for *compact* models of preferences. Some probabilistic models, like Bayesian networks or Markov random fields, fall in this category, as well as e.g., additive utilities [Fis67] and their generalisations. This focus on combinatorial objects also brought to light one difficulty with the utilitarian model: although it is often easy to compute the utilities or probabilities associated with two objects and compare them on such a basis, it appears to be often NP-hard to find optimal objects from a combinatorial set with numerical representations of preferences. Thus one other contribution of research in KRR is to provide preference representation languages where optimisation is computationally easy, like CP-nets [BBD⁺04].

These complex models of preferences have recently been studied from an ML perspective, both in an elicitation / active learning setting, and in a batch / passive learning setting. One particularity of these compact preference models is that they combine two elements: a structural element, indicating probabilistic or preferential interdependencies between the various features characterizing the objects of interest; and "local" preferences over small combinations of features. It is the structure learning phase which is often demanding, since finding the structure that best fits some data is often a hard combinatorial search problem. In contrast, finding the local preferences once the structure has been chosen is often easy.

The passive learning setting is particularly promising because of the vast dataset available in potential applications of preference learning in some decision aid systems like recommender systems or search engines. The possibility to learn Bayesian networks from data has been a key element for their early success in many applications. Note that in some applications, in particular in the study of biological systems, learning the structure, that is, the interdependencies between features, is interesting; in such applications, "black-box" models like deep neural networks seem less appropriate. This is also the case in decision-support systems where there is a need to *explain* the reasons justifying the computed ordering of possible decisions [BLM$^+$17].

At the frontier between learning and reasoning lies what could be named lazy preference learning: given a set of preference statements which do not specify a complete preference relation, one can infer new pairwise comparisons between objects, by assuming some properties the full, unknown preference relation. As a baseline, many settings in the models studied in KRR assume transitivity of preferences, but this alone does not usually induce many new comparisons. A common additional assumption, made by [BLM$^+$17], is that the preference relation can be represented with an additive utility function, and that the ordering over the domain of each feature is known. In [Wil06, Wil09, Wil14], richer classes of input preference statements are considered, and the assumption is made that the preference relation has some kind of (unknown) lexicographic structure.

Lastly, as mentioned above, analogical proportions can be used for predicting preferences [FH18, BPP18, BPPS19]. The general idea is that a preference between two items can be predicted if some analogical proportions hold that link their descriptions with the descriptions of other items for which preference relations are known.

5   Conclusion

The KRR and ML areas of AI have been developed independently to a large extent for several decades. It results that most of the researchers in one area are ignorant of what has been going on in the other area. The intended purpose of this joint work is to provide an inventory of the meeting points between KRR and ML lines of research. We have first reviewed some concerns that are shared by the two areas, maybe in different ways. Then we have surveyed various paradigms that are at the border between KRR and ML. Lastly, we have given an overview of different hybridizations of KRR and ML tools.

Let us emphasize that this is a work in progress. Subsections may be at this step unequally developed, and remain sketchy. The current list of references is certainly incomplete and unbalanced. The works covered may be old or recent, well-known as well as overlooked. However, at this step, we have absolutely no claim of completeness of any kind, not even of being fully up to date. Examples of topics not covered at all are numerous. They include reinforcement learning, argumentation and ML [AS08], ontology representation and learning [ASS15], in relation with the modeling of concepts, rough sets and rule extraction [Paw91, GZ00], or formal studies of data such as version space learning [Mit79] or logical analysis of data [BCH$^+$11]; see also [CLL$^+$13, Mir11].

Moreover, each topic covered in this paper is only outlined and would deserve to be discussed in further details. This paper is only a first step towards a more encompassing and extensive piece of work.

The aim of this paper is to help facilitating the understanding between researchers in the two areas, with a perspective of cross-fertilisation and mutual benefits. Still we should be aware that the mathematics of ML and the mathematics of KRR are quite different if we consider the main trends in each area. In ML the basic paradigm is a matter of approximating functions (which then calls for optimization). The mathematics of ML are close to those of signal processing and automatic control (as pointed out in [Le 19]), while KRR is dominated by logic and discrete mathematics, leading to an – at least apparent – opposition

between geometry and logic [Mal18][3]. But functions also underlie KRR, once one notices that a set of (fuzzy or weighted) rules is like an aggregation function [DP97], whose computation may take the form of a generalized matrix calculus ('generalized' in the sense that the operations are not necessarily restricted to sum and product). Let us also note that the convolution of functions (a key tool in signal processing) is no longer restricted to a linear, sum/product-based setting [MP19].

Besides, it seems difficult to envisage ML without KRR. For instance, it has been recently observed that

- the unsupervised learning of disentangled representations is fundamentally impossible without inductive biases on both the models and the data [LBL+18];

- local explanation methods for deep neural networks lack sensitivity to parameter values [AGGK18];

- when trained on one task, then trained on a second task, many machine learning models "forget" how to perform the first task [GMD+14].

Such states of fact might call for some cooperation in the long range between ML and KRR.

6 Acknowledgements

References

[AAA19]   AAAI. *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*. AAAI Press, 2019.

[AB18]   A. Adadi and M. Berrada. Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access*, 6:52138–52160, 2018.

[AC17]   A. Antonucci and G. Corani. The multilabel naive credal classifier. *Int. J. of Approximate Reasoning*, 83:320–336, 2017.

[ACM18]   J. M. Alonso, C. Castiello, and C. Mencar. A bibliometric analysis of the explainable artificial intelligence research field. In J. Medina, M. Ojeda-Aciego, J. L. V. Galdeano, D. A. Pelta, I. P. Cabrera, B. Bouchon-Meunier, and R. R. Yager, editors, *Proc. 17th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'18, Cádiz, June 11-15, Part I*, volume 853 of *Communications in Computer and Information Science*, pages 3–15. Springer, 2018.

[AFGG17]   A. M. Aung, Y. Fadila, R. Gondokaryono, and L. Gonzalez. Building robust deep neural networks for road sign detection. *CoRR*, abs/1712.09327, 2017.

[AGGK18]   J. Adebayo, J. Gilmer, I. J. Goodfellow, and B. Kim. Local explanation methods for deep neural networks lack sensitivity to parameter values. In *Workshop Track Proc. 6th Int. Conf. on Learning Representations, ICLR'18, Vancouver, April 30 - May 3*. OpenReview.net, 2018.

[AH15]   A. Abramé and D. Habet. AHMAXSAT : Description and evaluation of a branch and bound Max-SAT solver. *Journal on Satisfiability, Boolean Modeling and Computation*, 2015.

[AIS93]   R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases. In *Proc. SIGMOD Conf, Washington, DC, May 26-28*, pages 207–216. ACM Press, 1993.

---

[3]See also the conference https://www.youtube.com/watch?v=BpX890StRvs

[AKP10]   Z. Assaghir, M. Kaytoue, and H. Prade. A possibility theory-oriented discussion of conceptual pattern structures. In A. Deshpande and A. Hunter, editors, *Proc. 4th Int. Conf. (SUM'10)on Scalable Uncertainty Management Toulouse, Sept. 27-29*, volume 6379 of *LNCS*, pages 70–83. Springer, 2010.

[AKS16]   T. Ager, O. Kuzelka, and S. Schockaert. Inducing symbolic rules from entity embeddings using auto-encoders. In *NeSy@ HLAI*, 2016.

[ALA+17a] E. Angelino, N. Larus-Stone, D. Alabi, M. Seltzer, and C. Rudin. Learning certifiably optimal rule lists. In *Proc. 23rd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, Halifax, August 13 - 17*, pages 35–44. ACM, 2017.

[ALA+17b] E. Angelino, N. Larus-Stone, D. Alabi, M. Seltzer, and C. Rudin. Learning certifiably optimal rule lists for categorical data. *J. Mach. Learn. Res.*, 18:234:1–234:78, 2017.

[AMW19a]   S. Amizadeh, S. Matusevych, and M. Weimer. Learning to solve Circuit-SAT: An unsupervised differentiable approach. In *7th Int. Conf. on Learning Representations (ICLR'19), New Orleans, LA, May 6-9*. OpenReview.net, 2019.

[AMW19b]   S. Amizadeh, S. Matusevych, and M. Weimer. PDP: A general neural framework for learning constraint satisfaction solvers. *CoRR*, abs/1903.01969, 2019.

[ANCF19]   S. Anjomshoae, A. Najjar, D. Calvaresi, and K. Främling. Explainable agents and robots: Results from a systematic literature review. In *AAMAS*, pages 1078–1088, 2019.

[AP94]   A. Aamodt and E. Plaza. Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Commun.*, 7(1):39–59, 1994.

[AR90]   M. Ayel and M.-C. Rousset. *La Cohérence dans les Bases de Connaissances*. Cepadues, 1990.

[AS08]   L. Amgoud and M. Serrurier. Agents that argue and explain classifications. *Autonomous Agents and Multi-Agent Systems*, 16(2):187–209, 2008.

[ASS15]   R. Arp, B. Smith, and A. Spear. *Building Ontologies with Basic Formal Ontology*. MIT Press, 2015.

[Bar10]   C. Baral. *Knowledge Representation, Reasoning and Declarative Problem Solving*. Cambridge University Press, 2003, 2010.

[BBB+16]   J. Baget, S. Benferhat, Z. Bouraoui, M. Croitoru, M. Mugnier, O. Papini, S. Rocher, and K. Tabia. Inconsistency-tolerant query answering: Rationality properties and computational complexity analysis. In *Logics in Artificial Intelligence - 15th European Conference, JELIA 2016, Larnaca, Cyprus, November 9-11, 2016, Proceedings*, pages 64–80, 2016.

[BBD+04]   C. Boutilier, R. I. Brafman, C. Domshlak, H. H. Hoos, and D. Poole. CP-nets: a tool for representing and reasoning with conditional ceteris paribus preference statements. *Journal of Artificial Intelligence Research*, 21:135–191, 2004.

[BBL05]   F. Baader, S. Brandt, and C. Lutz. Pushing the EL envelope. In L. P. Kaelbling and A. Saffiotti, editors, *Proc. 19th Int. Joint Conf. on Artificial Intelligence (IJCAI'05), Edinburgh, July 30 - Aug. 5*, pages 364–369, 2005.

[BBV18]   M. Balunovic, P. Bielik, and M. T. Vechev. Learning to solve SMT formulas. In Bengio et al. [BWL+18], pages 10338–10349.

[BC17]   O. Biran and C. Cotton. Explanation and justification in machine learning: A survey. In *IJCAI-17 workshop on explainable AI (XAI)*, volume 8, page 1, 2017.

[BCD+18]   Q. Brabant, M. Couceiro, D. Dubois, H. Prade, and A. Rico. Extracting decision rules from qual-
           itative data via Sugeno utility functionals. In J. Medina, M. Ojeda-Aciego, J. L. V. Galdeano,
           D. A. Pelta, I. P. Cabrera, B. Bouchon-Meunier, and R. R. Yager, editors, *Proc. 17th Int. Conf. on
           Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'18),
           Cádiz, June 11-15, Part I*, volume 853 of *CCIS*, pages 253–265. Springer, 2018.

[BCH+11]   E. Boros, Y. Crama, P. L. Hammer, T. Ibaraki, A. Kogan, and K. Makino. Logical analysis of
           data: classification with justification. *Annals OR*, 188(1):33–61, 2011.

[BdGB+17] T. R. Besold, A. S. d'Avila Garcez, S. Bader, H. Bowman, P. M. Domingos, P. Hitzler, K. Kühn-
           berger, L. C. Lamb, D. Lowd, P. M. V. Lima, L. de Penning, G. Pinkas, H. Poon, and G. Za-
           verucha. Neural-symbolic learning and reasoning: A survey and interpretation. *CoRR*,
           abs/1711.03902, 2017.

[BDGP02]   S. Benferhat, D. Dubois, L. Garcia, and H. Prade. On the transformation between possibilistic
           logic bases and possibilistic causal networks. *Int. J. Approx. Reasoning*, 29(2):135–173, 2002.

[BDLP03]   S. Benferhat, D. Dubois, S. Lagrue, and H. Prade. A big-stepped probability approach for
           discovering default rules. *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems*,
           11(Supplement-1):1–14, 2003.

[BDP97]   S. Benferhat, D. Dubois, and H. Prade. Some syntactic approaches to the handling of inconsis-
           tent knowledge bases: A comparative study. Part 1: The flat case. *Studia Logica*, 58(1):17–45,
           1997.

[BDP99]   S. Benferhat, D. Dubois, and H. Prade. An overview of inconsistency-tolerant inferences in pri-
           oritized knowledge bases. In D. Dubois, H. Prade, and E. P. Klement, editors, *Fuzzy Sets, Logics
           and Reasoning about Knowledge*, pages 395–417. Springer Netherlands, Dordrecht, 1999.

[Bel02]   R. Belohlavek. *Fuzzy Relational Systems. Foundations and Principles*. Kluwer, 2002.

[BG91]   C. Balkenius and P. Gärdenfors. Nonmonotonic inferences in neural networks. In *Proc. 2nd
           Int. Conf. on Princip. of Knowl. Represent. and Reas. (KR'91). Cambridge, MA*, pages 32–39,
           1991.

[BGN14]   B. Babaki, T. Guns, and S. Nijssen. Constrained clustering using column generation. In *Inte-
           gration of AI and OR Techniques in Constraint Programming - Proc. 11th Int. Conf. CPAIOR'14,
           Cork, Ireland, May 19-23*, pages 438–454, 2014.

[BH98]   P. Besnard and A. Hunter. *Reasoning with Actual and Potential Contradictions*. Handbook
           of Defeasible Reasoning and Uncertainty Management Systems, Vol.2, (D. Gabbay and Ph.
           Smets, series editors. Kluwer Acad. Publ., 1998.

[BHD+18]   R. Bunel, M. J. Hausknecht, J. Devlin, R. Singh, and P. Kohli. Leveraging grammar and rein-
           forcement learning for neural program synthesis. In *6th Int. Conf. on Learning Representations
           (ICLR'18), Vancouver, April 30 - May 3, Conf. Track Proc.* OpenReview.net, 2018.

[BHKS98]   S.-E. Bornscheuer, S. Hölldobler, Y. Kalinke, and A. Strohmaier. Massively parallel reasoning.
           In W. Bibel and P. H. Schmitt, editors, *Automated Deduction - A Basis for Applications. Volume
           II: Systems and Implementation Techniques*, pages 291–321. Springer, 1998.

[BHO09]   C. Bessiere, E. Hebrard, and B. O'Sullivan. Minimising decision tree size as combinatorial
           optimisation. In I. P. Gent, editor, *Principles and Practice of Constraint Programming - CP
           2009, Proc. 15th Int. Conf. CP'09, Lisbon, Sept. 20-24*, volume 5732 of *LNCS*, pages 173–187.
           Springer, 2009.

[BKS18]   S. Bhatia, P. Kohli, and R. Singh. Neuro-symbolic program corrector for introductory pro-
           gramming assignments. In M. Chaudron, I. Crnkovic, M. Chechik, and M. Harman, editors,
           *Proce. 40th Int. Conf. on Software Engineering (ICSE'18), Gothenburg, Sweden, May 27 - June
           3*, pages 60–70. ACM, 2018.

[BL04]   R. Brachman and H. Levesque. *Knowledge Representation and Reasoning*. Elsevier, 2004. with a contribution of M. Pagnucco.

[BLM+17]   K. Belahcene, C. Labreuche, N. Maudet, V. Mousseau, and W. Ouerdane. Explaining robust additive utility models by sequences of preference swaps. *Theory and Decision*, 82(2):151–183, 2017.

[BLP18]   Y. Bengio, A. Lodi, and A. Prouvost. Machine learning for combinatorial optimization: a methodological tour d'horizon. *CoRR*, abs/1811.06128, 2018.

[BLR+19]   K. Bansal, S. M. Loos, M. N. Rabe, C. Szegedy, and S. Wilcox. HOList: An environment for machine learning of higher order logic theorem proving. In *Proc. 36th Int. Conf. on Machine Learning (ICML'19), 9-15 June, Long Beach, California*, pages 454–463, 2019.

[BLT+19]   R. Bunel, J. Lu, I. Turkaslan, P. H. S. Torr, P. Kohli, and M. P. Kumar. Branch and bound for piecewise linear neural network verification. *CoRR*, abs/1909.06588, 2019.

[BMT11]   G. Brewka, V. W. Marek, and M. Truszczynski, editors. *Nonmonotonic Reasoning. Essays Celebrating Its 30th Anniversary*, volume 31 of *Studies in Logic*. College Publication, 2011.

[BMV19]   M. Baader, M. Mirman, and M. T. Vechev. Universal approximation with certified networks. *CoRR*, abs/1909.13846, 2019.

[BPL+17]   I. Bello, H. Pham, Q. V. Le, M. Norouzi, and S. Bengio. Neural combinatorial optimization with reinforcement learning. In *5th International Conference on Learning Representations (ICLR'17), Toulon, Apr. 24-26, Workshop Track Proc.* OpenReview.net, 2017.

[BPP18]   M. Bounhas, M. Pirlot, and H. Prade. Predicting preferences by means of analogical proportions. In M. Cox, P. Funk, and S. Begum, editors, *Proc. 26th Int. Conf. on Case-Based Reason- ing Research (ICCBR'18), Stockholm, July 9-12*, volume 11156 of *LNCS*, pages 515–531. Springer, 2018.

[BPPS19]   M. Bounhas, M. Pirlot, H. Prade, and O. Sobrie. Comparison of analogy-based methods for predicting preferences. In N. B. Amor, B. Quost, and M. Theobald, editors, *Proc. 13th Int. Conf on Scalable Uncertainty Management (SUM'19), Compiègne, France, December 16-18*, volume 11940 of *LNCS*, pages 339–354. Springer, 2019.

[BPR17]   M. Bounhas, H. Prade, and G. Richard. Analogy-based classifiers for nominal or numerical data. *Int. J. Approx. Reasoning*, 91:36–55, 2017.

[BPT+00]   Y. Bastide, N. Pasquier, R. Taouil, G. Stumme, and L. Lakhal. Mining minimal non-redundant association rules using frequent closed itemsets. In J. W. Lloyd, V. Dahl, U. Furbach, M. Kerber, K. Lau, C. Palamidessi, L. M. Pereira, Y. Sagiv, and P. J. Stuckey, editors, *Computational Logic - CL 2000, Proc. 1st Int. Conf., London, 24-28 July*, volume 1861 of *LNCS*, pages 972–986. Springer, 2000.

[Bro92]   S. Bromberger, editor. *On What we Know we Don't Know: Explanation, Theory, Linguistics, and How Questions Shape Them*. University of Chicago Press, 1992.

[BS18]   Z. Bouraoui and S. Schockaert. Learning conceptual space representations of interrelated concepts. In *Proc. 27th Int. Joint Conf. on Artificial Intelligence*, pages 1760–1766, 2018.

[BS19]   Z. Bouraoui and S. Schockaert. Automated rule base completion as bayesian concept induction. In *Proc. 33rd AAAI Conference on Artificial Intelligence*, 2019.

[BSS11]   J. Błaszczyński, R. Słowiński, and M. Szelag. Sequential covering rule induction algorithm for variable consistency rough set approaches. *Information Sciences*, 181(5):987 – 1002, 2011.

[BSS+19]   T. Baluta, S. Shen, S. Shinde, K. S. Meel, and P. Saxena. Quantitative verification of neural networks and its security applications. In L. Cavallaro, J. Kinder, X. Wang, and J. Katz, editors, *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, CCS 2019, London, UK, November 11-15, 2019*, pages 1249–1264. ACM, 2019.

[BT12]   S. Benferhat and K. Tabia. Inference in possibilistic network classifiers under uncertain observations. *Ann. Math. Artif. Intell.*, 64(2-3):269–309, 2012.

[BTT+17]   R. Bunel, I. Turkaslan, P. H. S. Torr, P. Kohli, and M. P. Kumar. Piecewise linear neural network verification: A comparative study. *CoRR*, abs/1711.00455, 2017.

[BTT+18]   R. Bunel, I. Turkaslan, P. H. S. Torr, P. Kohli, and P. K. Mudigonda. A unified view of piecewise linear neural network verification. In *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems (NeurIPS'18), 3-8 Dec., Montréal*, pages 4795–4804, 2018.

[BUGD+13] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko. Translating embeddings for modeling multi-relational data. In *Advances in neural information processing systems*, pages 2787–2795, 2013.

[BWL+18]   S. Bengio, H. M. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors. *Advances in Neural Information Processing Systems 31: Annual Conf. on Neural Information Processing Systems (NeurIPS'18), Dec. 3-8, Montréal*, 2018.

[CAD+18]   A. Chakraborty, M. Alam, V. Dey, A. Chattopadhyay, and D. Mukhopadhyay. Adversarial attacks and defences: A survey. *CoRR*, abs/1810.00069, 2018.

[CAMO17]   A. Cornuéjols, S. Akkoyunlu, P.-A. Murena, and R. Olivier. Transfer learning by boosting projections from target to source. In *Conférence Francophone sur l'Apprentissage Automatique (CAP'17)*, 2017.

[Cas95]   J. Castro. Fuzzy logic controllers are universal approximators. *IEEE Trans. Systems, Man, and Cybernetics*, 25:629 – 635, 1995.

[CAZ12]   G. Corani, A. Antonucci, and M. Zaffalon. Bayesian networks with imprecise probabilities: Theory and application to classification. In *Data Mining: Foundations and Intelligent Paradigms*, pages 49–93. Springer, 2012.

[CBZ+19]   D. Chen, Y. Bai, W. Zhao, S. Ament, J. M. Gregoire, and C. P. Gomes. Deep reasoning networks: Thinking fast and slow. *CoRR*, abs/1906.00855, 2019.

[CC16]   W. Carnielli and M. Coniglio. *Paraconsistent Logic: Consistency, Contradiction and Negation*. Springer, 2016.

[CD16]   I. Couso and D. Dubois. Belief revision and the EM algorithm. In J. P. Carvalho, M. Lesot, U. Kaymak, S. M. Vieira, B. Bouchon-Meunier, and R. R. Yager, editors, *Information Processing and Management of Uncertainty in Knowledge-Based Systems - Proc. 16th International Conference, IPMU 2016, Part II*, volume 611 of *Communications in Computer and Information Science*, pages 279–290. Springer, 2016.

[CD18]   I. Couso and D. Dubois. A general framework for maximizing likelihood under incomplete data. *Int. J. of Approximate Reasoning*, 93:238–260, 2018.

[CDH+16]   X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In *Advances in Neural Information Processing Systems*, pages 2172–2180, 2016.

[CDH17]   I. Couso, D. Dubois, and E. Hüllermeier. Maximum likelihood estimation and coarse data. In *Proc 11th Int. Conf. on Scalable Uncertainty Management (SUM'17)*, volume 10564 of *LNCS*, pages 3–16. Springer, 2017.

[CDL⁺05]   D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, and R. Rosati. Dl-lite: Tractable description logics for ontologies. In M. M. Veloso and S. Kambhampati, editors, *Proc. 20th National Conf. on Artificial Intelligence (AAAI'05), July 9-13, Pittsburgh*, pages 602–607. AAAI Press / The MIT Press, 2005.

[CDL19]   L. Charnay, J. Dibie, and S. Loiseau. Validation and explanation. In P. Marquis, O. Papini, and H. Prade, editors, *A Guided Tour of Artificial Intelligence Research. Vol. 1: Knowledge Representation, Reasoning and Learning*. Springer, 2019.

[CFTR16]   N. Courty, R. Flamary, D. Tuia, and A. Rakotomamonjy. Optimal transport for domain adaptation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 39(9):1853–1865, 2016.

[CHPR17]   M. Couceiro, N. Hug, H. Prade, and G. Richard. Analogy-preserving functions: A way to extend boolean samples. In *Proc. 26th Int. Joint Conf. on Artificial Intelligence, (IJCAI'17), Melbourne, Aug. 19-25*, pages 1575– 1581, 2017.

[CHPR18]   M. Couceiro, N. Hug, H. Prade, and G. Richard. Behavior of analogical inference w.r.t. Boolean functions. In *Proc. 27th Int. Joint Conf. on Artificial Intelligence, (IJCAI'18), Stockholm, July. 13-19,* pages 2057–2063, 2018.

[CHWW12]   W. Cheng, E. Hüllermeier, W. Waegeman, and V. Welker. Label ranking with partial abstention based on thresholded probabilistic models. In P. L. Bartlett, F. C. N. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25: 26th Annual Conf. on Neural Information Processing Systems 2012. Proc. of a meeting held December 3-6, Lake Tahoe, Nevada*, pages 2510–2518, 2012.

[CJSU19]   K. Chvalovský, J. Jakubuv, M. Suda, and J. Urban. ENIGMA-NG: efficient neural and gradient-boosted inference guidance for E. In P. Fontaine, editor, *Automated Deduction - CADE'19 - Proc. 27th Int. Conf. on Automated Deduction, Natal, Brazil, Aug. 27-30*, volume 11716 of *LNCS*, pages 197–215. Springer, 2019.

[CJSW18]   P. Chalasani, S. Jha, A. Sadagopan, and X. Wu. Adversarial learning and explainability in structured datasets. *CoRR*, abs/1810.06583, 2018.

[CKL⁺19]   L. Cardelli, M. Kwiatkowska, L. Laurenti, N. Paoletti, A. Patane, and M. Wicker. Statistical guarantees for the robustness of bayesian neural networks. In S. Kraus, editor, *Proc. 28th Int. Joint Conf. on Artificial Intelligence (IJCAI'19), Macao, Aug. 10-16*, pages 5693–5700. ijcai.org, 2019.

[CKN20]   A. Cornuejols, F. Koriche, and R. Nock. Statistical computational learning. In P. Marquis, O. Papini, and H. Prade, editors, *A Guided Tour of Artificial Intelligence Research. Vol. 1 Knowledge Representation, Reasoning and Learning,* pages 341–388. Springer-Verlag, 2020.

[CLL⁺13]   I. Chikalov, V. V. Lozin, I. Lozina, M. Moshkov, H. S. Nguyen, A. Skowron, and B. Zielosko. *Three Approaches to Data Analysis - Test Theory, Rough Sets and Logical Analysis of Data*, volume 41 of *Intelligent Systems Reference Library*. Springer, 2013.

[CM09]   M. Chein and M.-L. Mugnier. *Graph-based Knowledge Representation: Computational Foundations of Conceptual Graphs*. Springer, Princeton, N.J., 2009.

[COC⁺12]   Z. L. Cherfi, L. Oukhellou, E. Côme, T. Denœux, and P. Aknin. Partially supervised independent factor analysis using soft labels elicited from multiple experts: Application to railway track circuit diagnosis. *Soft Computing*, 16(5):741–754, 2012.

[Coz00]   F. G. Cozman. Credal networks. *Artif. Intell.*, 120(2):199–233, 2000.

[Coz05]   F. G. Cozman. Graphical models for imprecise probabilities. *Int. J. of Approximate Reasoning*, 39:167–184, 2005.

[Coz20]   F. G. Cozman. Languages for probabilistic modeling over structured domains. In P. Marquis, O. Papini, and H. Prade, editors, *A Guided Tour of Artificial Intelligence Research. Vol. 2 Artificial Intelligence Algorithms*, pages 247–283. Springer-Verlag, 2020.

[CS12]   J. Cid-Sueiro. Proper losses for learning from partial labels. In *Advances in neural information processing systems*, pages 1565–1573, 2012.

[CS16]   I. Couso and L. Sánchez. Machine learning models, epistemic set-valued data and generalized loss functions: An encompassing approach. *Information Sciences*, 358:129–150, 2016.

[CST11]   T. Cour, B. Sapp, and B. Taskar. Learning from partial labels. *Journal of Machine Learning Research*, 12(May):1501–1536, 2011.

[CSZ06]   O. Chapelle, B. Schölkopf, and A. Zien. *Semi-Supervised Learning*. MIT Press, 2006.

[CV20]   A. Cornuejols and C. Vrain. Designing algorithms for machine learning and data mining. In P. Marquis, O. Papini, and H. Prade, editors, *A Guided Tour of Artificial Intelligence Research. Vol. 2 Artificial Intelligence Algorithms*, pages 339–410. Springer-Verlag, 2020.

[CY19]   Z. Chen and Z. Yang. Graph neural reasoning may fail in certifying boolean unsatisfiability. *CoRR*, abs/1909.11588, 2019.

[dAN94]   F. d'Alché-Buc, V. Andrés, and J. Nadal. Rule extraction with fuzzy neural network. *Int. J. Neural Syst.*, 5(1):1–11, 1994.

[DBH18]   F. K. Dosilovic, M. Brcic, and N. Hlupic. Explainable artificial intelligence: A survey. In *MIPRO*, pages 210–215, 2018.

[DBW18]   A. Dittadi, T. Bolander, and O. Winther. Learning to plan from raw data in grid-based games. In D. D. Lee, A. Steen, and T. Walsh, editors, *GCAI-2018, 4th Global Conference on Artificial Intelligence, Luxembourg, September 18-21, 2018*, volume 55 of *EPiC Series in Computing*, pages 54–67. EasyChair, 2018.

[DCZ04]   G. De Cooman and M. Zaffalon. Updating beliefs with incomplete observations. *Artificial Intelligence*, 159(1-2):75–125, 2004.

[DDV17]   T. Dao, K. Duong, and C. Vrain. Constrained clustering by constraint programming. *Artif. Intell.*, 244:70–94, 2017.

[De 08]   L. De Raedt. *Logical and Relational Learning*. Springer, 2008.

[Dem67]   A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38:325–339, 1967.

[Den95]   T. Denœux. A k-nearest neighbor classification rule based on Dempster-Shafer theory. *IEEE Trans. on Systems, Man and Cybernetics*, 25(05):804–813, 1995.

[Den13]   T. Denœux. Maximum likelihood estimation from uncertain data in the belief function framework. *IEEE Trans. on Knowledge and Data Engineering*, 25(1):119–130, 2013.

[Den19]   T. Denœux. Logistic regression, neural networks and Dempster-Shafer theory: A new perspective. *Knowl.-Based Syst.*, 176:54–67, 2019.

[DF36]   B. De Finetti. La logique des probabilités. In *Congrès International de Philosophie Scientifique*, pages 1–9, Paris, France, 1936. Hermann et Cie.

[DFG$^+$19] T. Dreossi, D. J. Fremont, S. Ghosh, E. Kim, H. Ravanbakhsh, M. Vazquez-Chanlatte, and S. A. Seshia. VERIFAI: A toolkit for the design and analysis of artificial intelligence-based systems. *CoRR*, abs/1902.04245, 2019.

[DFKM08]   L. De Raedt, P. Frasconi, K. Kersting, and S. Muggleton, editors. *Probabilistic Inductive Logic Programming - Theory and Applications*, volume 4911 of *Lecture Notes in Computer Science*. Springer, 2008.

[DGFK18]   K. Dvijotham, M. Garnelo, A. Fawzi, and P. Kohli. Verification of deep probabilistic models. *CoRR*, abs/1812.02795, 2018.

[dGGL⁺19] A. S. d'Avila Garcez, M. Gori, L. C. Lamb, L. Serafini, M. Spranger, and S. N. Tran. Neural-symbolic computing: An effective methodology for principled integration of machine learning and reasoning. *CoRR*, abs/1905.06088, 2019.

[dGL03]   A. S. d'Avila Garcez and L. C. Lamb. Reasoning about time and knowledge in neural symbolic learning systems. In S. Thrun, L. K. Saul, and B. Schölkopf, editors, *Advances in Neural Information Processing Systems 16 (NIPS 2003)*, pages 921–928. MIT Press, 2003.

[dGLG03]   A. S. d'Avila Garcez, L. C. Lamb, and D. M. Gabbay. Neural-symbolic intuitionistic reasoning. In A. Abraham, M. Köppen, and K. Franke, editors, *Proceedings of the Third International Conference on Hybrid Intelligent Systems*, volume 105 of *Frontiers in Artificial Intelligence and Applications*, pages 399–408. IOS Press, 2003.

[dGLG06]   A. S. d'Avila Garcez, L. C. Lamb, and D. M. Gabbay. Connectionist computations of intuitionistic reasoning. *Theoretical Computer Science*, 358(1):34–55, 2006.

[dGLG07]   A. S. d'Avila Garcez, L. C. Lamb, and D. M. Gabbay. Connectionist modal logic: Representing modalities in neural networks. *Theoretical Computer Science*, 371(1-2):34–53, 2007.

[dGLG09]   A. S. d'Avila Garcez, L. C. Lamb, and D. M. Gabbay. *Neural-Symbolic Cognitive Reasoning*. Cognitive Technologies. Springer, 2009.

[DGMR12]   M. Diligenti, M. Gori, M. Maggini, and L. Rigutini. Bridging logic and kernel machines. *Machine Learning*, 86(1):57–88, 2012.

[DGN08]   L. De Raedt, T. Guns, and S. Nijssen. Constraint programming for itemset mining. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Las Vegas, Nevada, USA, August 24-27, 2008*, pages 204–212, 2008.

[DGN10]   L. De Raedt, T. Guns, and S. Nijssen. Constraint programming for data mining and machine learning. In *Proc. 24th AAAI Conf. on Artificial Intelligence, (AAAI'10), Atlanta, July 11-15*, 2010.

[dGZ99]   A. S. d'Avila Garcez and G. Zaverucha. The connectionist inductive learning and logic programming system. *Applied Intelligence*, 11(1):59–77, 1999.

[DH07]   D. Dubois and E. Hüllermeier. Comparing probability measures using possibility theory: A notion of relative peakedness. *Int. J. Approx. Reasoning*, 45(2):364–385, 2007.

[DHP06]   D. Dubois, E. Hüllermeier, and H. Prade. A systematic approach to the assessment of fuzzy association rules. *Data Min. Knowl. Discov.*, 13(2):167–192, 2006.

[DKC⁺03]   K. Duan, S. S. Keerthi, W. Chu, S. K. Shevade, and A. N. Poo. Multi-category classification by soft-max combination of binary classifiers. In T. Windeatt and F. Roli, editors, *Proceedings of the 4th International Workshop on Multiple Classifier Systems (MCS 2003)*, volume 2709 of *Lecture Notes in Computer Science*, pages 125–134. Springer, 2003.

[DKNP16]   L. De Raedt, K. Kersting, S. Natarajan, and D. Poole. *Statistical Relational Artificial Intelligence: Logic, Probability, and Computation*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers, 2016.

[DKS19] T. Denœux, O. Kanjanatarakul, and S. Sriboonchitta. A new evidential k-nearest neighbor rule based on contextual discounting with partially supervised learning. *International Journal of Approximate Reasoning*, 113:287–302, 2019.

[DKT07] L. De Raedt, A. Kimmig, and H. Toivonen. Problog: A probabilistic prolog and its application in link discovery. In *IJCAI 2007, Proceedings of the 20th International Joint Conference on Artificial Intelligence, Hyderabad, India, January 6-12, 2007*, pages 2462–2467, 2007.

[DL01] S. Dzeroski and N. Lavrac, editors. *Relational data mining*. Springer, 2001.

[DL18] T. Denœux and S. Li. Frequency-calibrated belief functions: Review and new insights. *Int. J. of Approximate Reasoning*, 92:232–254, 2018.

[DLP94] D. Dubois, J. Lang, and H. Prade. Possibilistic logic. In D. Gabbay, C. Hogger, J. Robinson, and D. Nute, editors, *Handbook of Logic in Artificial Intelligence and Logic Programming, Vol. 3*, pages 439–513. Oxford University Press, 1994.

[DLR77] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *J. of the Royal Statistical Society, Series B*, 39(1):1–38, 1977.

[DM02] A. Darwiche and P. Marquis. A knowledge compilation map. *Journal of Artificial Intelligence Research*, 17:229–264, 2002.

[DML+19] H. Dong, J. Mao, T. Lin, C. Wang, L. Li, and D. Zhou. Neural logic machines. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019.

[DP97] D. Dubois and H. Prade. Fuzzy criteria and fuzzy rules in subjective evaluation – A general discussion. In *Proc. 5th Eur. Cong. Intel. Techn. Soft Comput. (EUFIT'97), Aachen, Vol. 1, 975-979*, 1997.

[DP12] D. Dubois and H. Prade. Possibility theory and formal concept analysis: Characterizing independent sub-contexts. *Fuzzy Sets and Systems*, 196:4–16, 2012.

[DP19] D. Dubois and H. Prade. Towards a reconciliation between reasoning and learning - A position paper. In N. B. Amor, B. Quost, and M. Theobald, editors, *Proc. 13th Int. Conf. on Scalable Uncertainty Management (SUM'19), Compiègne, Dec. 16-18*, volume 11940 of *LNCS*, pages 153–168. Springer, 2019.

[DPR14] D. Dubois, H. Prade, and A. Rico. The logical encoding of Sugeno integrals. *Fuzzy Sets and Systems*, 241:61–75, 2014.

[DPR16] D. Dubois, H. Prade, and G. Richard. Multiple-valued extensions of analogical proportions. *Fuzzy Sets and Systems*, 292:193–202, 2016.

[DPS05] D. Dubois, H. Prade, and T. Sudkamp. On the representation, measurement, and discovery of fuzzy associations. *IEEE Trans. Fuzzy Systems*, 13(2):250–262, 2005.

[DPS17] D. Dubois, H. Prade, and S. Schockaert. Generalized possibilistic logic: Foundations and applications to qualitative reasoning about uncertainty. *Artif. Intell.*, 252:139–174, 2017.

[DRR16] T. Demeester, T. Rocktäschel, and S. Riedel. Lifted rule injection for relation embeddings. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1389–1399, 2016.

[DRS10] I. Davidson, S. S. Ravi, and L. Shamis. A sat-based framework for efficient constrained clustering. In *Proc. SIAM Int. Conf. on Data Mining (SDM'10), April 29 - May 1, Columbus, Ohio*, pages 94–105, 2010.

[DS15]   J. Derrac and S. Schockaert. Inducing semantic relations from conceptual spaces: a data-driven approach to plausible reasoning. *Artificial Intelligence*, pages 74–105, 2015.

[DSdG17]   I. Donadello, L. Serafini, and A. S. d'Avila Garcez. Logic tensor networks for semantic image interpretation. In C. Sierra, editor, *Proc. of the 26th Int. Joint Conf. on Artificial Intelligence (IJCAI'17)*, pages 1596–1602. ijcai.org, 2017.

[DSG⁺18]   K. Dvijotham, R. Stanforth, S. Gowal, T. A. Mann, and P. Kohli. A dual approach to scalable verification of deep networks. In A. Globerson and R. Silva, editors, *Proceedings of the Thirty-Fourth Conference on Uncertainty in Artificial Intelligence, UAI 2018, Monterey, California, USA, August 6-10, 2018*, pages 550–559. AUAI Press, 2018.

[DSG⁺19]   K. D. Dvijotham, R. Stanforth, S. Gowal, C. Qin, S. De, and P. Kohli. Efficient neural network verification with exactness characterization. In A. Globerson and R. Silva, editors, *Proceedings of the Thirty-Fifth Conference on Uncertainty in Artificial Intelligence, UAI 2019, Tel Aviv, Israel, July 22-25, 2019*, page 164. AUAI Press, 2019.

[DVDD16]   T. Dao, C. Vrain, K. Duong, and I. Davidson. A framework for actionable clustering using constraint programming. In *ECAI 2016 - 22nd European Conference on Artificial Intelligence, 29 Aug.-2 Sept. 2016, The Hague*, pages 453–461, 2016.

[DXYZ18]   W. Dai, Q. Xu, Y. Yu, and Z. Zhou. Tunneling neural perception and logic reasoning through abductive learning. *CoRR*, abs/1802.01173, 2018.

[DZ01]   T. Denœux and L. M. Zouhal. Handling possibilistic labels in pattern classification using evidential reasoning. *Fuzzy Sets and Systems*, 122(3):47–62, 2001.

[DŽ04]   S. Džeroski and B. Ženko. Is combining classifiers with stacking better than selecting the best one? *Machine learning*, 54(3):255–273, 2004.

[EEF⁺18]   K. Eykholt, I. Evtimov, E. Fernandes, B. Li, A. Rahmati, C. Xiao, A. Prakash, T. Kohno, and D. Song. Robust physical-world attacks on deep learning visual classification. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*, pages 1625–1634. IEEE Computer Society, 2018.

[EG18a]   R. Evans and E. Grefenstette. Learning explanatory rules from noisy data. *J. Artif. Intell. Res.*, 61:1–64, 2018.

[EG18b]   R. Evans and E. Grefenstette. Learning explanatory rules from noisy data (extended abstract). In Lang [Lan18], pages 5598–5602.

[Erk09]   K. Erk. Representing words as regions in vector space. In *Proc. 13th Conf. on Computational Natural Language Learning*, pages 57–65, 2009.

[ES15]   R. Eldan and O. Shamir. The power of depth for feedforward neural networks. *CoRR*, abs/1512.03965, 2015.

[ESA⁺18]   R. Evans, D. Saxton, D. Amos, P. Kohli, and E. Grefenstette. Can neural networks understand logical entailment? In *6th Int. Conf. on Learning Representations (ICLR'18), Vancouver, April 30 - May 3, Conf. Track Proc.*, 2018.

[FB12]   A. Flint and M. B. Blaschko. Perceptron learning of SAT. In P. L. Bartlett, F. C. N. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012. Proceedings of a meeting held December 3-6, 2012, Lake Tahoe, Nevada, United States*, pages 2780–2788, 2012.

[FBBR05]   D. Fierens, H. Blockeel, M. Bruynooghe, and J. Ramon. Logical bayesian networks and their relation to other probabilistic logical models. In *Inductive Logic Programming, Proc. 15th Int. Conf. ILP'05, Bonn, Germany, Aug. 10-13*, pages 121–135, 2005.

[FBD⁺19] M. Fischer, M. Balunovic, D. Drachsler-Cohen, T. Gehr, C. Zhang, and M. T. Vechev. DL2: training and querying neural networks with logic. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, pages 1931–1941, 2019.

[FBI⁺19] S. G. Finlayson, J. D. Bowers, J. Ito, J. L. Zittrain, A. L. Beam, and I. S. Kohane. Adversarial attacks on medical machine learning. *Science*, 363(6433):1287–1289, 2019.

[FH10] J. Fürnkranz and E. Hüllermeier, editors. *Preference Learning*. Springer, 2010.

[FH18] M. A. Fahandar and E. Hüllermeier. Learning to rank based on analogical reasoning. In *Proc. 32th Nat Conf. on Artificial Intelligence (AAAI'18), New Orleans, Feb. 2-7*, 2018.

[FHR⁺14] J. Fürnkranz, E. Hüllermeier, C. Rudin, R. Slowinski, and S. Sanner. Preference learning (dagstuhl seminar 14101). *Dagstuhl Reports*, 4(3):1–27, 2014.

[Fis67] P. C. Fishburn. Interdependence and additivity in multivariate, unidimensional expected utility theory. *International Economic Review*, 8(3):pp. 335–342, 1967.

[FKH⁺20] S. Ferré, M. Kaytoue, M. Huchard, S. O. Kuznetsov, and A. Napoli. Formal concept analysis: From knowledge discovery to knowledge processing. In P. Marquis, O. Papini, and H. Prade, editors, *A Guided Tour of Artificial Intelligence Research. Vol. 2 Artificial Intelligence Algorithms*, pages 411–445. Springer-Verlag, 2020.

[FP89] H. Farreny and H. Prade. Positive and negative explanations of uncertain reasoning in the framework of possibility theory. In *Proc. 5th Conference on Uncertainty in Artificial Intelligence (UAI'89), Windsor, ON, Aug. 18-20*, pages 95–101, 1989. Available in CoRR, abs/1304.1502, 2013; Expanded version: Explications de raisonnements dans l'incertain), Revue d'Intelligence Artificielle, 4(2), 43-75,1990.

[FR04] S. Ferré and O. Ridoux. Introduction to logical information systems. *Information Process. & Manag.*, 40(3):383–419, 2004.

[FZdG14] M. V. M. França, G. Zaverucha, and A. S. d'Avila Garcez. Fast relational learning using bottom clause propositionalization with artificial neural networks. *Machine Learning*, 94(1):81–104, 2014.

[Gär91] P. Gärdenfors. Nonmonotonic inference, expectations, and neural networks. In R. Kruse and P. Siegel, editors, *Proc. Europ. Conf. on Symbolic and Quantitative Approaches to Reasoning and Uncertainty (ECSQA), Marseille, Oct. 15-17*, volume 548 of *LNCS*, pages 12–27. Springer, 1991.

[Gär00] P. Gärdenfors. *Conceptual Spaces: The Geometry of Thought*. MIT Press, 2000.

[GBC16] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016.

[GBG02] A. S. d. Garcez, K. B. Broda, and D. M. Gabbay. *Neural-symbolic learning systems: foundations and applications*. Springer Science & Business Media, 2002.

[GBP18] A. Gupta, G. Boleda, and S. Padó. Instantiation. *CoRR*, abs/1808.01662, 2018.

[GBY⁺18] L. H. Gilpin, D. Bau, B. Z. Yuan, A. Bajwa, M. Specter, and L. Kagal. Explaining explanations: An overview of interpretability of machine learning. In F. Bonchi, F. J. Provost, T. Eliassi-Rad, W. Wang, C. Cattuto, and R.Ghani, editors, *Proc. 5th IEEE Int. Conf. on Data Science and Advanced Analytics (DSAA'18), Turin, Oct. 1-3*, pages 80–89. IEEE, 2018.

[GD86] J. L. Guigues and V. Duquenne. Familles minimales d'implications informatives résultant d'un tableau de données binaires. *Mathématiques et Sciences Humaines*, 95:5–18, 1986.

[GD18]   R. Guillaume and D. Dubois. A maximum likelihood approach to inference under coarse data based on minimax regret. In S. Destercke, T. Denœux, M. A. Gil, P. Grzegorzewski, and O. Hryniewicz, editors, *Uncertainty Modelling in Data Science, SMPS 2018*, volume 832 of *Advances in Intelligent Systems and Computing*, pages 99–106. Springer, 2018.

[GIS06]   S. Greco, M. Inuiguchi, and R. Slowinski. Fuzzy rough sets and multiple-premise gradual decision rules. *Int. J. Approx. Reasoning*, 41(2):179–211, 2006.

[GK01]   B. Ganter and S. O. Kuznetsov. Pattern structures and their projections. In H. S. Delugach and G. Stumme, editors, *Proc. 9th Int. Conf. on Conceptual Structures (ICCS'01)*, volume 2120 of *LNCS*, pages 129–142. Springer, 2001.

[GKPB18]   D. Gopinath, G. Katz, C. S. Pasareanu, and C. W. Barrett. Deepsafe: A data-driven approach for assessing robustness of neural networks. In Lahiri and Wang [LW18], pages 3–19.

[GL05]   G. P. G. Leng, Th. M. McGinnity. An approach for on-line extraction of fuzzy rules using a self-organising fuzzy neural network. *Fuzzy Sets and Systems*, 150:211—-243, 2005.

[GL10]   M. Grabisch and C. Labreuche. A decade of application of the Choquet and Sugeno integrals in multi-criteria decision aid. *Annals of Oper. Res.*, 175:247–286, 2010.

[GM19]   B. Ghosh and K. S. Meel. IMLI: an incremental framework for maxsat-based learning of interpretable classification rules. In V. Conitzer, G. K. Hadfield, and S. Vallor, editors, *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, AIES 2019, Honolulu, HI, USA, January 27-28, 2019*, pages 203–210. ACM, 2019.

[GMD+14]   I. J. Goodfellow, M. Mirza, X. Da, A. C. Courville, and Y. Bengio. An empirical investigation of catastrophic forgetting in gradient-based neural networks. In *Conf. Track Proc. 2nd Int. Conf. on Learning Representations (ICLR'14), Banff, April 14-16*, 2014.

[GMD+18]   T. Gehr, M. Mirman, D. Drachsler-Cohen, P. Tsankov, S. Chaudhuri, and M. T. Vechev. AI2: safety and robustness certification of neural networks with abstract interpretation. In *2018 IEEE Symposium on Security and Privacy, SP 2018, Proceedings, 21-23 May 2018, San Francisco, California, USA*, pages 3–18. IEEE Computer Society, 2018.

[GMR+19]   R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi. A survey of methods for explaining black box models. *ACM Comput. Surv.*, 51(5):93:1–93:42, 2019.

[GMS04]   S. Greco, B. Matarazzo, and R. Slowinski. Axiomatic characterization of a general utility function and its particular cases in terms of conjoint measurement and rough-set decision rules. *Europ.J. of Operational Research*, 158(2):271–292, 2004.

[GP14]   C. Grozea and M. Popescu. Can machine learning learn a decision oracle for NP problems? A test on SAT. *Fundam. Inform.*, 131(3-4):441–450, 2014.

[GRKC09]   Y. Grandvalet, A. Rakotomamonjy, J. Keshet, and S. Canu. Support vector machines with a reject option. In *Advances in neural information processing systems*, pages 537–544, 2009.

[GS18]   V. Gutiérrez-Basulto and S. Schockaert. From knowledge graph embedding to ontology embedding? an analysis of the compatibility between vector space representations and rules. In *Proc. of the 16th Int. Conf. on Principles of Knowledge Representation and Reasoning*, pages 379–388, 2018.

[GSR+17]   J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl. Neural message passing for quantum chemistry. In D. Precup and Y. W. Teh, editors, *Proc. of the 34th Int. Conf. on Machine Learning, (ICML'17), Sydney, 6-11 Aug.*, volume 70 of *Proceedings of Machine Learning Research*, pages 1263–1272. PMLR, 2017.

[GSS14]   I. J. Goodfellow, J. Shlens, and C. Szegedy. Explaining and harnessing adversarial examples, 2014.

[GT07]   L. Getoor and B. Taskar, editors. *Introduction to Statistical Relational Learning*. Adaptive Computation and Machine Learning. MIT Press, 2007.

[GTFA19]   L. H. Gilpin, C. Testart, N. Fruchter, and J. Adebayo. Explaining explanations to society. *CoRR*, abs/1901.06560, 2019.

[GvLB+17] I. Guyon, U. von Luxburg, S. Bengio, H. M. Wallach, R. Fergus, S. V. N. Vishwanathan, and R. Garnett, editors. *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA*, 2017.

[GVV98]   A. Gammerman, V. Vovk, and V. Vapnik. Learning by transduction. In *Proc. 14th Conf. on Uncertainty in AI*, pages 148–155. Morgan Kaufmann, 1998.

[GW98]   B. Ganter and R. Wille. *Formal Concept Analysis: Mathematical Foundations*. Springer-Verlag, 1998.

[GZ00]   J. W. Grzymala-Busse and W. Ziarko. Data mining and rough set theory. *Commun. ACM*, 43(4):108–109, 2000.

[Ha97]   T. M. Ha. The optimum class-selective rejection rule. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(6):608–615, 1997.

[Hal17]   J. Y. Halpern. *Reasoning about Uncertainty*. MIT Press, 2017.

[HC15]   E. Hüllermeier and W. Cheng. Superset learning based on generalized loss minimization. In *Machine Learning and Knowledge Discovery in Databases - Proc. Eur. Conf., ECML PKDD 2015, Part II*, volume 9285 of *Lecture Notes in Computer Science*, pages 260–275. Springer, 2015.

[HCKB16]   F. Hill, K. Cho, A. Korhonen, and Y. Bengio. Learning to understand phrases by embedding the dictionary. *Transactions of the Association for Computational Linguistics*, 4:17–30, 2016.

[HCR84]   D. W. Hasling, W. J. Clancey, and G. Rennels. Strategic explanations for a diagnostic consultation system. *Int. J. of Man-Machine Studies*, 20(1):3–19, 1984.

[HCS+16]   I. Hubara, M. Courbariaux, D. Soudry, R. El-Yaniv, and Y. Bengio. Binarized neural networks. In Lee et al. [LSvL+16], pages 4107–4115.

[HDC19]   E. Hüllermeier, S. Destercke, and I. Couso. Learning from imprecise data: Adjustments of optimistic and pessimistic variants. In *Scalable Uncertainty Management - 13th International Conference, SUM 2019, Compiègne, France, December 16-18, 2019, Proceedings*, pages 266–279, 2019.

[HDP02]   E. Hüllermeier, D. Dubois, and H. Prade. Model adaptation in possibilistic instance-based reasoning. *IEEE Trans. Fuzzy Systems*, 10(3):333–339, 2002.

[HDSS19]   D. Huang, P. Dhariwal, D. Song, and I. Sutskever. Gamepad: A learning environment for theorem proving. In *7th Int. Conf. on Learning Representations (ICLR'19, New Orleans, LA, May 6-9, 2019*. OpenReview.net, 2019.

[Hea19]   D. Heaven. Why deep-learning AIs are so easy to fool. *Nature*, 574(7777):163, 2019.

[HFMV03]   J. Y. Halpern, R. Fagin, Y. Moses, and M. Y. Vardi. *Reasoning about Knowledge*. MIT Press, 1995 & 2003.

[HH78]   P. Hájek and P. Havránek. *Mechanising Hypothesis Formation - Mathematical Foundations for a General Theory*. Springer Verlag, 1978.

[HHL11]   F. Hutter, H. H. Hoos, and K. Leyton-Brown. Sequential model-based optimization for general algorithm configuration. In C. A. C. Coello, editor, *Learning and Intelligent Optimization - 5th International Conference, LION 5, Rome, Italy, January 17-21, 2011. Selected Papers*, volume 6683 of *LNCS*, pages 507–523. Springer, 2011.

[HHLS09]   F. Hutter, H. H. Hoos, K. Leyton-Brown, and T. Stützle. ParamILS: An automatic algorithm configuration framework. *J. Artif. Intell. Res.*, 36:267–306, 2009.

[HK17a]   R. R. Hoffman and G. Klein. Explaining explanation, part 1: Theoretical foundations. *IEEE Intelligent Systems*, 32(3):68–73, 2017.

[HK17b]   R. R. Hoffman and G. Klein. Explaining explanation, part 1: Theoretical foundations. *IEEE Intelligent Systems*, 32(3):68–73, 2017.

[HKS99]   S. Hölldobler, Y. Kalinke, and H. Störr. Approximating the semantics of logic programs by recurrent neural networks. *Applied Intelligence*, 11(1):45–58, 1999.

[HKWW17] X. Huang, M. Kwiatkowska, S. Wang, and M. Wu. Safety verification of deep neural networks. In Majumdar and Kuncak [MK17], pages 3–29.

[HL17]   P. Hohenecker and T. Lukasiewicz. Deep learning for ontology reasoning. *CoRR*, abs/1705.10342, 2017.

[HMK17]   R. R. Hoffman, S. T. Mueller, and G. Klein. Explaining explanation, part 2: Empirical foundations. *IEEE Intelligent Systems*, 32(4):78–86, 2017.

[HMKL18]   R. R. Hoffman, S. T. Mueller, G. Klein, and J. Litman. Metrics for explainable AI: challenges and prospects. *CoRR*, abs/1812.04608, 2018.

[HML$^+$16]   Z. Hu, X. Ma, Z. Liu, E. H. Hovy, and E. P. Xing. Harnessing deep neural networks with logic rules. In *Proc. 54th Annual Meeting of the Association for Computational Linguistics*, 2016.

[HMM$^+$18]   R. R. Hoffman, T. Miller, S. T. Mueller, G. Klein, and W. J. Clancey. Explaining explanation, part 4: A deep dive on deep nets. *IEEE Intelligent Systems*, 33(3):87–95, 2018.

[HMP$^+$17]   I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner. β-VAE: Learning basic visual concepts with a constrained variational framework. In *International Conference on Learning Representations*, volume 3, 2017.

[Hoo18]   J. N. Hooker, editor. *Principles and Practice of Constraint Programming - Proc. 24th Int. Conf. CP 2018, Lille, Aug. 27-31*, volume 11008 of *Lecture Notes in Computer Science*. Springer, 2018.

[Hoy04]   P. O. Hoyer. Non-negative matrix factorization with sparseness constraints. *Journal of Machine Learning Research*, 5:1457–1469, 2004.

[HP05]   J. Y. Halpern and J. Pearl. Causes and explanations: A structural-model approach. part II: explanations. *The British Journal for the Philosophy of Science*, 56(4):889–911, 2005.

[HR76]   L. Hyafil and R. L. Rivest. Constructing optimal binary decision trees is np-complete. *Information Processing Letters*, 5(1):15–17, 1976.

[HR91]   D. F. Heitjan and D. B. Rubin. Ignorability and coarse data. *Ann. Statist.*, 19(4):2244–2253, 1991.

[HRS19]   X. Hu, C. Rudin, and M. I. Seltzer. Optimal sparse decision trees. In *NeurIPS*, 2019.

[Hua19]   D. Huang. On learning to prove. *CoRR*, abs/1904.11099, 2019.

[Hül14]   E. Hüllermeier. Learning from imprecise and fuzzy observations: Data disambiguation through generalized loss minimization. *International Journal of Approximate Reasoning*, 55(7):1519–1534, 2014.

[HW09]   S. Haim and T. Walsh. Restart strategy selection using machine learning techniques. In O. Kull-mann, editor, *Theory and Applications of Satisfiability Testing - SAT 2009, 12th International Conference, SAT 2009, Swansea, UK, June 30 - July 3, 2009. Proceedings*, volume 5584 of *LNCS*, pages 312–325. Springer, 2009.

[HXHL14]   F. Hutter, L. Xu, H. H. Hoos, and K. Leyton-Brown. Algorithm runtime prediction: Methods & evaluation. *Artif. Intell.*, 206:79–111, 2014.

[ICL17]   ICLR. *5th International Conference on Learning Representations (ICLR'17), Toulon, April 24-26, 2017, Conf. Track Proc.* OpenReview.net, 2017.

[INM19a]   A. Ignatiev, N. Narodytska, and J. Marques-Silva. Abduction-based explanations for machine learning models. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019* [AAA19], pages 1511–1519.

[INM19b]   A. Ignatiev, N. Narodytska, and J. Marques-Silva. On validating, repairing and refining heuris-tic ML explanations. *CoRR*, abs/1907.02509, 2019.

[IPNM18]   A. Ignatiev, F. Pereira, N. Narodytska, and J. Marques-Silva. A SAT-based approach to learn explainable decision sets. In D. Galmiche, S. Schulz, and R. Sebastiani, editors, *Automated Reasoning - Proc. 9th Int. Joint Conf. IJCAR'18, Held as Part of the Federated Logic Conference, FloC 2018, Oxford, UK, July 14-17*, volume 10900 of *LNCS*, pages 627–645. Springer, 2018.

[ISA⁺16]   G. Irving, C. Szegedy, A. A. Alemi, N. Eén, F. Chollet, and J. Urban. DeepMath - deep sequence models for premise selection. In Lee et al. [LSvL⁺16], pages 2235–2243.

[Jae05]   M. Jaeger. Ignorability in statistical and probabilistic inference. *J. Artif. Intell. Res.*, 24:889–917, 2005.

[Jan18]   M. Janota. Towards generalization in QBF solving via machine learning. In *Proc. 32nd AAAI Conf. on Artificial Intelligence, (AAAI-18), New Orleans, Feb. 2-7*, pages 6607–6614, 2018.

[Jef83]   R. Jeffrey. *The logic of decision*. 2nd ed. Chicago University Press, 1983.

[JL19]   M. Janota and I. Lynce, editors. *Theory and Applications of Satisfiability Testing - SAT 2019 - Proc. 22nd Int. Conf. SAT'19, Lisbon, July 9-12*, volume 11628 of *LNCS*. Springer, 2019.

[JS93]   J. Jang and C. Sun. Functional equivalence between radial basis function networks and fuzzy inference systems. *IEEE Trans. Neural Networks*, 4(1):156–159, 1993.

[JS16]   S. Jameel and S. Schockaert. Entity embeddings with conceptual subspaces as a basis for plausible reasoning. In *Proc. 22nd Europ. Conf. on Artificial Intelligence (ECAI'16), 29 Aug.-2 Sept. 2016, The Hague*, pages 1353–1361, 2016.

[JS17]   S. Jameel and S. Schockaert. Modeling context words as regions: An ordinal regression ap-proach to word embedding. In *Proc. 21st Conf. on Computational Natural Language Learning*, pages 123–133, 2017.

[JSS17]   S. Jabbour, L. Sais, and Y. Salhi. Mining top-k motifs with a sat-based framework. *Artif. Intell.*, 244:30–47, 2017.

[Kas87]   G. Kassel. The use of deep knowledge to improve explanation capabilities of rule-based expert systems. In H. Balzert, G. Heyer, and R. Lutze, editors, *Expertensysteme '87: Konzepte und Werkzeuge, Tagung I/1987 des German Chapter of the ACM am 7. und 8.4.1987 in Nürnberg*, volume 28 of *Berichte des German Chapter of the ACM*, pages 315–326. Teubner, 1987.

[KBC10]   M. Khiari, P. Boizumault, and B. Crémilleux. Constraint programming for mining n-ary patterns. In *Principles and Practice of Constraint Programming - Proc. CP 16th Int. Conf. CP 2010, St. Andrews, Scotland, Sept. 6-10*, pages 552–567, 2010.

[KBD+17]   G. Katz, C. W. Barrett, D. L. Dill, K. Julian, and M. J. Kochenderfer. Reluplex: An efficient SMT solver for verifying deep neural networks. In Majumdar and Kuncak [MK17], pages 97–117.

[KCS17]   C. Kaliszyk, F. Chollet, and C. Szegedy. Holstep: A machine learning dataset for higher-order logic theorem proving. In *5th International Conference on Learning Representations(ICLR'17), Toulon, April 24-26, 2017, Conf. Track Proc.* [ICL17].

[KDS15]   O. Kuzelka, J. Davis, and S. Schockaert. Encoding Markov logic networks in possibilistic logic. In M. Meila and T. Heskes, editors, *Proc. 31st Conf. on Uncertainty in Artificial Intelligence (UAI'15), July 12-16, Amsterdam*, pages 454–463. AUAI Press, 2015.

[KDZ+17]   E. B. Khalil, H. Dai, Y. Zhang, B. Dilkina, and L. Song. Learning combinatorial optimization algorithms over graphs. In Guyon et al. [GvLB+17], pages 6348–6358.

[KFF15]   A. Karpathy and L. Fei-Fei. Deep visual-semantic alignments for generating image descriptions. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3128–3137, 2015.

[KHI+19]   G. Katz, D. A. Huang, D. Ibeling, K. Julian, C. Lazarus, R. Lim, P. Shah, S. Thakoor, H. Wu, A. Zeljic, D. L. Dill, M. J. Kochenderfer, and C. W. Barrett. The marabou framework for verification and analysis of deep neural networks. In I. Dillig and S. Tasiran, editors, *Computer Aided Verification - 31st International Conference, CAV 2019, New York City, NY, USA, July 15-18, 2019, Proceedings, Part I*, volume 11561 of *LNCS*, pages 443–452. Springer, 2019.

[KKK+17]   N. Kaur, G. Kunapuli, T. Khot, K. Kersting, W. Cohen, and S. Natarajan. Relational restricted boltzmann machines: A probabilistic logic learning approach. In *Inductive Logic Programming - 27th International Conference, ILP 2017, Orléans, France, September 4-6, 2017, Revised Selected Papers*, pages 94–111, 2017.

[Kle18]   G. Klein. Explaining explanation, part 3: The causal landscape. *IEEE Intelligent Systems*, 33(2):83–88, 2018.

[KLL+17]   A. Kemmar, Y. Lebbah, S. Loudni, P. Boizumault, and T. Charnois. Prefix-projection global constraint and top-k approach for sequential pattern mining. *Constraints*, 22(2):265–306, 2017.

[KLM90]   S. Kraus, D. Lehmann, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44:167–207, 1990.

[KP18]   S. M. Kazemi and D. Poole. Simple embedding for link prediction in knowledge graphs. In *Advances in Neural Information Processing Systems 31: Annual Conf. on Neural Information Processing Systems 2018, NeurIPS 2018, 3-8 December 2018, Montréal, Canada.*, pages 4289–4300, 2018.

[KRD+17]   C. Kuo, S. S. Ravi, T. Dao, C. Vrain, and I. Davidson. A framework for minimal clustering modification via constraint programming. In *Proc. 31st AAAI Conf. on Artificial Intelligence, February 4-9, 2017, San Francisco*, pages 1389–1395, 2017.

[KSD16]   O. Kanjanatarakul, S. Sriboonchitta, and T. Denœux. Statistical estimation and prediction using belief functions: principles and application to some econometric models. *Int. J. of Approximate Reasoning*, 72:71–94, 2016.

[KSS+16]   B. Krishnapuram, M. Shah, A. J. Smola, C. C. Aggarwal, D. Shen, and R. Rastogi, editors. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*. ACM, 2016.

[KU14] C. Kaliszyk and J. Urban. Learning-assisted automated reasoning with Flyspeck. *J. Autom. Reasoning*, 53(2):173–213, 2014.

[KU15] C. Kaliszyk and J. Urban. Learning-assisted theorem proving with millions of lemmas. *J. Symb. Comput.*, 69:109–128, 2015.

[KUMO18] C. Kaliszyk, J. Urban, H. Michalewski, and M. Olsák. Reinforcement learning of theorem proving. In Bengio et al. [BWL+18], pages 8836–8847.

[KUV14] C. Kaliszyk, J. Urban, and J. Vyskocil. Machine learner for automated reasoning 0.4 and 0.5. In S. Schulz, L. de Moura, and B. Konev, editors, *4th Workshop on Practical Aspects of Automated Reasoning, PAAR@IJCAR 2014, Vienna, 2014*, volume 31 of *EPiC Series in Computing*, pages 60–66. EasyChair, 2014.

[Kwi19] M. Z. Kwiatkowska. Safety verification for deep neural networks with provable guarantees (invited paper). In W. Fokkink and R. van Glabbeek, editors, *30th International Conference on Concurrency Theory, CONCUR 2019, August 27-30, 2019, Amsterdam, the Netherlands*, volume 140 of *LIPIcs*, pages 1:1–1:5. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019.

[KXHL16] A. R. KhudaBukhsh, L. Xu, H. H. Hoos, and K. Leyton-Brown. SATenstein: automatically building local search SAT solvers from components. *Artif. Intell.*, 232:20–42, 2016.

[Lab11] C. Labreuche. A general framework for explaining the results of a multi-attribute preference model. *Artif. Intell.*, 175(7-8):1410–1448, 2011.

[Lan18] J. Lang, editor. *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden*. ijcai.org, 2018.

[LB87] H. J. Levesque and R. J. Brachman. Expressiveness and tractability in knowledge representation and reasoning. *Computational Intelligence*, 3:78–93, 1987.

[LBH15] Y. LeCun, Y. Bengio, and G. E. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[LBL16] H. Lakkaraju, S. H. Bach, and J. Leskovec. Interpretable decision sets: A joint framework for description and prediction. In Krishnapuram et al. [KSS+16], pages 1675–1684.

[LBL+18] F. Locatello, S. Bauer, M. Lucic, S. Gelly, B. Schölkopf, and O. Bachem. Challenging common assumptions in the unsupervised learning of disentangled representations. *CoRR*, abs/1811.12359, 2018.

[LBS19] N. Li, Z. Bouraoui, and S. Schockaert. Ontology completion using graph convolutional networks. In *The Semantic Web - ISWC 2019 - 18th International Semantic Web Conference, Auckland, New Zealand, October 26-30, 2019, Proceedings, Part I*, pages 435–452, 2019.

[LCK18] Z. Li, Q. Chen, and V. Koltun. Combinatorial optimization with graph convolutional networks and guided tree search. In Bengio et al. [BWL+18], pages 537–546.

[LCW+15] C. Luo, S. Cai, W. Wu, Z. Jie, and K. Su. CCLS: an efficient local search algorithm for weighted maximum satisfiability. *IEEE Transactions on Computers*, 64(7):1830–1843, 2015.

[LD14] L. Liu and T. Dietterich. Learnability of the superset label learning problem. In *International Conference on Machine Learning*, pages 1629–1637, 2014.

[LDG91] N. Lavrac, S. Dzeroski, and M. Grobelnik. Learning nonrecursive definitions of relations with LINUS. In *Machine Learning - EWSL-91, European Working Session on Learning, Porto, Portugal, March 6-8, 1991, Proceedings*, pages 265–281, 1991.

[Le 19] Y. Le Cun. *Quand la Machine Apprend. La révolution des neurones artificiels et de l'apprentissage profond*. Odile Jacob, 2019.

[LF02]   N. Lachiche and P. A. Flach. 1bc2: A true first-order bayesian classifier. In *Inductive Logic Programming, 12th International Conference, ILP 2002, Sydney, Australia, July 9-11, 2002. Revised Papers*, pages 133–148, 2002.

[LGPC16a] J. H. Liang, V. Ganesh, P. Poupart, and K. Czarnecki. Exponential recency weighted average branching heuristic for SAT solvers. In D. Schuurmans and M. P. Wellman, editors, *Proc. 30th AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix*, pages 3434–3440. AAAI Press, 2016.

[LGPC16b] J. H. Liang, V. Ganesh, P. Poupart, and K. Czarnecki. Learning rate based branching heuristic for SAT solvers. In N. Creignou and D. L. Berre, editors, *Theory and Applications of Satisfiability Testing - SAT 2016 - Proc. 19th Int. Conf., Bordeaux, July 5-8*, volume 9710 of *LNCS*, pages 123–140. Springer, 2016.

[LISK17]   S. M. Loos, G. Irving, C. Szegedy, and C. Kaliszyk. Deep network guided proof search. In T. Eiter and D. Sands, editors, *LPAR-21, 21st International Conference on Logic for Programming, Artificial Intelligence and Reasoning, Maun, Botswana, May 7-12, 2017*, volume 46 of *EPiC Series in Computing*, pages 85–105. EasyChair, 2017.

[LL17]   S. M. Lundberg and S. Lee. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, pages 4765–4774, 2017.

[LNPR18]   J. Lieber, E. Nauer, H. Prade, and G. Richard. Making the best of cases by approximation, interpolation and extrapolation. In M. T. Cox, P. Funk, and S. Begum, editors, *Proc. 26th Int. Conf. on Case-Based Reasoning (ICCBR'18), Stockholm, July 9-12*, volume 11156 of *LNCS*, pages 580–596. Springer, 2018.

[LNPT18]   F. Leofante, N. Narodytska, L. Pulina, and A. Tacchella. Automated verification of neural networks: Advances, challenges and perspectives. *CoRR*, abs/1805.09938, 2018.

[LOM+18]   J. H. Liang, C. Oh, M. Mathew, C. Thomas, C. Li, and V. Ganesh. Machine learning-based restart policy for CDCL SAT solvers. In O. Beyersdorff and C. M. Wintersteiger, editors, *Theory and Applications of Satisfiability Testing - SAT 2018 - Proc. 21st Inte. Conf. SAT'18, Held as Part of the Federated Logic Conference, FloC 2018, Oxford, July 9-12,*, volume 10929 of *LNCS*, pages 94–110. Springer, 2018.

[LPNC+17] D. Lopez-Paz, R. Nishihara, S. Chintala, B. Scholkopf, and L. Bottou. Discovering causal signals in images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6979–6987, 2017.

[LR19]   R. J. Little and D. B. Rubin. *Statistical Analysis with Missing Data*, volume 793. Wiley, 2019.

[LRD16]   C. Lian, S. Ruan, and T. Denœux. Dissimilarity metric learning in the belief function framework. *IEEE Transactions on Fuzzy Systems*, 24(6):1555–1564, 2016.

[LRS18]   G. Lederman, M. N. Rabe, and S. A. Seshia. Learning heuristics for automated reasoning through deep reinforcement learning. *CoRR*, abs/1807.08058, 2018.

[LS99]   D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788, 1999.

[LSvL+16] D. D. Lee, M. Sugiyama, U. von Luxburg, I. Guyon, and R. Garnett, editors. *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, 2016.

[LTC17]   M. Law, N. Thome, and M. Cord. Learning a distance metric from relative comparisons between quadruplets of images. *Int. J. of Computer Vision*, 121(1):65–94, 2017.

[LW18]   S. K. Lahiri and C. Wang, editors. *Automated Technology for Verification and Analysis - Proc. 16th Int. Symp. ATVA(18, Los Angeles, Oct. 7-10*, volume 11138 of *LNCS*. Springer, 2018.

[LYH18]   N. Liu, H. Yang, and X. Hu. Adversarial detection with model interpretation. In Y. Guo and F. Farooq, editors, *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, August 19-23, 2018*, pages 1803–1811. ACM, 2018.

[LZSJ19]   X. Lin, H. Zhu, R. Samanta, and S. Jagannathan. ART: abstraction refinement-guided training for provably correct neural networks. *CoRR*, abs/1907.10662, 2019.

[MA75]   E. H. Mamdani and S. Assilian. An experiment in linguistic synthesis with a fuzzy logic controller. *Int. J. of Man-Machine Studies*, 7(1):1 – 13, 1975.

[Mal18]   S. Mallat. *Sciences des Données et Apprentissage en Grande Dimension*. Leçons Inaugurales du Collège de France. Fayard, Paris, 2018.

[Mar15]   P. Marquis. Compile! In B. Bonet and S. Koenig, editors, *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI 2015)*, pages 4112–4118. AAAI Press, 2015.

[MB10]   C. Marsala and B. Bouchon-Meunier. Quality of measures for attribute selection in fuzzy decision trees. In *FUZZ-IEEE 2010, IEEE International Conference on Fuzzy Systems, Barcelona, Spain, 18-23 July, 2010, Proceedings*, pages 1–8. IEEE, 2010.

[MBD08]   L. Miclet, S. Bayoudh, and A. Delhay. Analogical dissimilarity: definition, algorithms and two experiments in machine learning. *JAIR, 32*, pages 793–824, 2008.

[MBRR18]   P. Minervini, M. Bosnjak, T. Rocktäschel, and S. Riedel. Towards neural theorem proving at scale. *CoRR*, abs/1807.08204, 2018.

[MD94]   S. Muggleton and L. De Raedt. Inductive logic programming: Theory and methods. *J. Log. Program.*, 19/20:629–679, 1994.

[MDK+18a] R. Manhaeve, S. Dumancic, A. Kimmig, T. Demeester, and L. De Raedt. Deepproblog: Neural probabilistic logic programming. In *Advances in Neural Information Processing Systems*, pages 3749–3759, 2018.

[MDK+18b] R. Manhaeve, S. Dumancic, A. Kimmig, T. Demeester, and L. De Raedt. DeepProbLog: Neural probabilistic logic programming. In Bengio et al. [BWL+18], pages 3753–3763.

[MDNS08]   N. Messai, M. Devignes, A. Napoli, and M. Smaïl-Tabbone. Many-valued concept lattices for conceptual clustering and information retrieval. In M. Ghallab, C. D. Spyropoulos, N. Fakotakis, and N. M. Avouris, editors, *Proc. 18th Europ. Conf. on Artificial Intelligence (ECAI'08), Patras, July 21-25*, volume 178 of *Frontiers in Artificial Intelligence and Applications*, pages 127–131. IOS Press, 2008.

[MGK+19]   J. Mao, C. Gan, P. Kohli, J. B. Tenenbaum, and J. Wu. The neuro-symbolic concept learner: Interpreting scenes, words, and sentences from natural supervision. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019.

[MGV18]   M. Mirman, T. Gehr, and M. T. Vechev. Differentiable abstract interpretation for provably robust neural networks. In J. G. Dy and A. Krause, editors, *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, volume 80 of *Proceedings of Machine Learning Research*, pages 3575–3583. PMLR, 2018.

[MHC+19]   S. T. Mueller, R. R. Hoffman, W. J. Clancey, A. Emrey, and G. Klein. Explanation in human-AI systems: A literature meta-review, synopsis of key ideas and publications, and bibliography for explainable AI. *CoRR*, abs/1902.01876, 2019.

[Mil19a]   T. Miller. "But why?" understanding explainable artificial intelligence. *ACM Crossroads*, 25(3):20–25, 2019.

[Mil19b]   T. Miller. Explanation in artificial intelligence: Insights from the social sciences. *Artif. Intell.*, 267:1–38, 2019.

[Mir11]   B. Mirkin. *Core Concepts in Data Analysis: Summarization, Correlation, Visualization*. Springer, 2011.

[Mit79]   T. Mitchell. *Version spaces: An approach to concept learning*. PhD thesis, Stanford University, 1979.

[MK10]   M. Mueller and S. Kramer. Integer linear programming models for constrained clustering. In *Discovery Science - Proc. 13th Int. Conf. DS'10, Canberra, Oct. 6-8*, pages 159–173, 2010.

[MK17]   R. Majumdar and V. Kuncak, editors. *Computer Aided Verification - 29th Int. Conf. CAV'17, Heidelberg, July 24-28, 2017, Proceedings, Part I*, volume 10426 of *LNCS*. Springer, 2017.

[MM18]   D. Malioutov and K. S. Meel. MLIC: A maxsat-based framework for learning interpretable classification rules. In Hooker [Hoo18], pages 312–327.

[MML14]   R. Martins, V. M. Manquinho, and I. Lynce. Open-wbo: A modular maxsat solver. In C. Sinz and U. Egly, editors, *Proceedings of the 17th International Conference on Theory and Applications of Satisfiability Testing (SAT 2014)*, volume 8561 of *Lecture Notes in Computer Science*, pages 438–445. Springer, 2014.

[MP09]   L. Miclet and H. Prade. Handling analogical proportions in classical logic and fuzzy logics settings. In *Proc. 10th Eur. Conf. on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'09)*, pages 638–650. Springer, LNCS 5590, 2009.

[MP19]   V. Molek and I. Perfilieva. Scale-space theory, F-transform kernels and CNN realization. In *Advances in Computational Intelligence - Proc. 15th Int. Work-Conf. on Artificial Neural Networks, IWANN 2019, Gran Canaria, June 12-14, Part II*, pages 38–48, 2019.

[MPLM08]   P. J. Matos, J. Planes, F. Letombe, and J. Marques-Silva. A MAX-SAT algorithm portfolio. In M. Ghallab, C. D. Spyropoulos, N. Fakotakis, and N. M. Avouris, editors, *ECAI 2008 - 18th European Conference on Artificial Intelligence, Patras, Greece, July 21-25, 2008, Proceedings*, volume 178 of *Frontiers in Artificial Intelligence and Applications*, pages 911–912. IOS Press, 2008.

[MPP19]   P. Marquis, O. Papini, and H. Prade, editors. *A Guided Tour of Artificial Intelligence Research. Vol. 1 Knowledge representation, reasoning and learning. Vol. 2 Artificial intelligence algorithms. Vol. 3 Interfaces and applications of artificial intelligence*. Springer, 2019.

[MRT18]   M. Mohri, A. Rostamizadeh, and A. Talwalkar. *Foundations of Machine Learning. Second edition*. MIT Press, 2018.

[MRW19]   B. D. Mittelstadt, C. Russell, and S. Wachter. Explaining explanations in AI. In *FAT*, pages 279–288, 2019.

[MSM18]   G. Montavon, W. Samek, and K. Müller. Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, 73:1–15, 2018.

[MSZ+18]   S. H. Muggleton, U. Schmid, C. Zeller, A. Tamaddoni-Nezhad, and T. R. Besold. Ultra-strong machine learning: comprehensibility of programs learned with ILP. *Machine Learning*, 107(7):1119–1140, 2018.

[Mug95]   S. Muggleton. Inverse entailment and Progol. *New Generation Comput.*, 13(3-4):245–286, 1995.

[Nar18]   N. Narodytska. Formal analysis of deep binarized neural networks. In Lang [Lan18], pages 5692–5696.

[NF07]   S. Nijssen and É. Fromont. Mining optimal decision trees from itemset lattices. In P. Berkhin, R. Caruana, and X. Wu, editors, *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Jose, California, USA, August 12-15, 2007*, pages 530–539. ACM, 2007.

[NF10]   S. Nijssen and É. Fromont. Optimal constraint-based decision tree induction from itemset lattices. *Data Min. Knowl. Discov.*, 21(1):9–51, 2010.

[Ngu78]   H. T. Nguyen. On random sets and belief functions. *J. of Mathematical Analysis and Applications*, 65:531–542, 1978.

[NH10]   V. Nair and G. E. Hinton. Rectified linear units improve restricted boltzmann machines. In J. Fürnkranz and T. Joachims, editors, *Proceedings of the 27th International Conference on Machine Learning (ICML-10), June 21-24, 2010, Haifa, Israel*, pages 807–814. Omnipress, 2010.

[Nij08]   S. Nijssen. Bayes optimal classification for decision trees. In W. W. Cohen, A. McCallum, and S. T. Roweis, editors, *Machine Learning, Proceedings of the Twenty-Fifth International Conference (ICML 2008), Helsinki, Finland, June 5-9, 2008*, volume 307 of *ACM International Conference Proceeding Series*, pages 696–703. ACM, 2008.

[NIPM18]   N. Narodytska, A. Ignatiev, F. Pereira, and J. Marques-Silva. Learning optimal decision trees with SAT. In Lang [Lan18], pages 1362–1368.

[NKR+18]   N. Narodytska, S. P. Kasiviswanathan, L. Ryzhyk, M. Sagiv, and T. Walsh. Verifying properties of binarized deep neural networks. In *Proc. 32nd AAAI Conf. on Artificial Intelligence, (AAAI-18), New Orleans, Louisiana, Feb. 2-7, 2018*, pages 6615–6624, 2018.

[NLP10]   J. Nin, A. Laurent, and P. Poncelet. Speed up gradual rule mining from stream data! A b-tree and owa-based approach. *J. Intell. Inf. Syst.*, 35(3):447–463, 2010.

[NSM+19]   N. Narodytska, A. A. Shrotri, K. S. Meel, A. Ignatiev, and J. Marques-Silva. Assessing heuristic machine learning explanations with model counting. In Janota and Lynce [JL19], pages 267–278.

[OZL+17]   A. Ouali, A. Zimmermann, S. Loudni, Y. Lebbah, B. Crémilleux, P. Boizumault, and L. Loukil. Integer linear programming for pattern set mining; with an application to tiling. In *Advances in Knowledge Discovery and Data Mining - 21st Pacific-Asia Conference, PAKDD 2017, Jeju, South Korea, May 23-26, 2017, Proceedings, Part II*, pages 286–299, 2017.

[PAL+19]   M. O. R. Prates, P. H. C. Avelar, H. Lemos, L. C. Lamb, and M. Y. Vardi. Learning to solve NP-complete problems: A graph neural network for decision TSP. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019* [AAA19], pages 4731–4738.

[Paw91]   Z. Pawlak. *Rough Sets. Theoretical Aspects of. Reasoning about Data*. Kluwer Acad. Publ., Dordrecht, 1991.

[PBTL99]   N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal. Efficient mining of association rules using closed itemset lattices. *Inf. Syst.*, 24:25–46, 1999.

[PC19]   G. Pinkas and S. Cohen. High-order networks that learn to satisfy logic constraints. *Journal of Applied Logics - IfCoLog Journal of Logics and their Applications*, 6(4):653–694, 2019.

[Pea88]   J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

[Pea09]   J. Pearl. *Causality*. Cambridge university press, 2009.

[Ped98]   W. Pedrycz. Conditional fuzzy clustering in the design of radial basis function neural networks. *IEEE Trans. Neural Networks*, 9(4):601–612, 1998.

[PGJ16]   J. Pearl, M. Glymour, and N. P. Jewell. *Causal inference in statistics: A primer*. John Wiley & Sons, 2016.

[Pin91]   G. Pinkas. Symmetric neural networks and propositional logic satisfiability. *Neural Computation*, 3(2):282–291, 1991.

[Pin95]   G. Pinkas. Reasoning, nonmonotonicity and learning in connectionist networks that capture propositional knowledge. *Artificial Intelligence*, 77(2):203 – 247, 1995.

[Plo70]   G. Plotkin. A note on inductive generalization. In *Machine Intelligence*, volume 5, pages 153–163. Edinburgh University Press, 1970.

[PLR+19]   A. Paliwal, S. M. Loos, M. N. Rabe, K. Bansal, and C. Szegedy. Graph representations for higher-order logic and theorem proving. *CoRR*, abs/1905.10006, 2019.

[PM18]   J. Pearl and D. Mackenzie. *The Book of Why: The New Science of Cause and Effect*. Basic Books, 2018.

[PMS+17]   E. Parisotto, A. Mohamed, R. Singh, L. Li, D. Zhou, and P. Kohli. Neuro-symbolic program synthesis. In *5th Int. Conf. on Learning Representations, ICLR'17, Toulon, France, April 24-26, Conf. Track Proc.* [ICL17].

[PPW18]   R. B. Palm, U. Paquet, and O. Winther. Recurrent relational networks. In Bengio et al. [BWL+18], pages 3372–3382.

[PR13]   H. Prade and G. Richard. From analogical proportion to logical proportions. *Logica Universalis*, 7(4):441–505, 2013.

[PR18a]   P. Panda and K. Roy. Explainable learning: Implicit generative modelling during training for adversarial robustness. *CoRR*, abs/1807.02188, 2018.

[PR18b]   H. Prade and G. Richard. Analogical proportions: From equality to inequality. *Int. J. Approx. Reasoning*, 101:234–254, 2018.

[Pra16]   H. Prade. Reasoning with data - A new challenge for AI? In S. Schockaert and P. Senellart, editors, *Proc. 10th Int. Conf. on Scalable Uncertainty Management(SUM'16), Nice, Sept. 21-23*, volume 9858 of *LNCS*, pages 274–288. Springer, 2016.

[PRS09]   H. Prade, A. Rico, and M. Serrurier. Elicitation of Sugeno integrals: A version space learning perspective. In J. Rauch, Z. W. Ras, P. Berka, and T. Elomaa, editors, *Proc. 18th Int. Symp. on Foundations of Intelligent Systems (ISMIS'09), Prague, Sept. 14-17*, volume 5722 of *LNCS*, pages 392–401. Springer, 2009.

[PRSR09]   H. Prade, A. Rico, M. Serrurier, and E. Raufaste. Elicitating Sugeno integrals:Methodology and a case study. In C. Sossai and G. Chemello, editors, *Proc. 10th Eur. Conf. on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'09), Verona, July 1-3*, volume 5590 of *LNCS*, pages 712–723. Springer, 2009.

[PT10]   L. Pulina and A. Tacchella. An abstraction-refinement approach to verification of artificial neural networks. In T. Touili, B. Cook, and P. B. Jackson, editors, *Computer Aided Verification, 22nd International Conference, CAV 2010, Edinburgh, UK, July 15-19, 2010. Proceedings*, volume 6174 of *LNCS*, pages 243–257. Springer, 2010.

[QDL17]   B. Quost, T. Denœux, and S. Li. Parametric classification with soft labels using the eviden-
           tial EM algorithm: linear discriminant analysis versus logistic regression. *Advances in Data
           Analysis and Classification*, 11(4):659–690, Dec 2017.

[QDO⁺19]   C. Qin, K. D. Dvijotham, B. O'Donoghue, R. Bunel, R. Stanforth, S. Gowal, J. Uesato,
           G. Swirszcz, and P. Kohli. Verification of non-linear specifications for neural networks. In
           *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA,
           May 6-9, 2019*. OpenReview.net, 2019.

[QMD11]   B. Quost, M.-H. Masson, and T. Denœux. Classifier fusion in the Dempster-Shafer framework
           using optimized t-norm based combination rules. *Int. J. of Approximate Reasoning*, 52(3):353–
           374, 2011.

[Qui96]   J. R. Quinlan. Learning first-order definitions of functions. *CoRR*, cs.AI/9610102, 1996.

[RD06]   M. Richardson and P. M. Domingos. Markov logic networks. *Machine Learning*, 62(1-2):107–
           136, 2006.

[RD18]   A. S. Ross and F. Doshi-Velez. Improving the adversarial robustness and interpretability of deep
           neural networks by regularizing their input gradients. In *Proc. 32nd AAAI Conf. on Artificial
           Intelligence, (AAAI-18), New Orleans, Feb. 2-7*, pages 1660–1669, 2018.

[RHK18]   W. Ruan, X. Huang, and M. Kwiatkowska. Reachability analysis of deep neural networks with
           provable guarantees. In Lang [Lan18], pages 2651–2659.

[Rog94]   G. Rogova. Combining the results of several neural network classifiers. *Neural Networks*,
           7(5):777–781, 1994.

[Ros58]   F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization
           in the brain. *Psychological Review*, 65(6):386–408, 1958.

[RR16]   T. Rocktäschel and S. Riedel. Learning knowledge base inference with neural theorem provers.
           In J. Pujara, T. Rocktäschel, D. Chen, and S. Singh, editors, *Proc. 5th Workshop on Automated
           Knowledge Base Construction, AKBC@NAACL-HLT 2016, San Diego, CA, USA, June 17, 2016*,
           pages 45–50. The Association for Computer Linguistics, 2016.

[RR17]   T. Rocktäschel and S. Riedel. End-to-end differentiable proving. In Guyon et al. [GvLB⁺17],
           pages 3788–3800.

[RS87]   M. Rousset and B. Safar. Negative and positive explanations in expert. *Applied Artificial
           Intelligence*, 1(1):25–38, 1987.

[RSG16]   M. T. Ribeiro, S. Singh, and C. Guestrin. "why should I trust you?": Explaining the predictions
           of any classifier. In Krishnapuram et al. [KSS⁺16], pages 1135–1144.

[RSG18]   M. T. Ribeiro, S. Singh, and C. Guestrin. Anchors: High-precision model-agnostic explana-
           tions. In S. A. McIlraith and K. Q. Weinberger, editors, *Proc. 32nd AAAI Conference on Artificial
           Intelligence (AAAI-18), New Orleans, Feb. 2-7*, pages 1527–1535, 2018.

[Rud19]   C. Rudin. Stop explaining black box machine learning models for high stakes decisions and
           use interpretable models instead. *Nature Machine Intelligence*, 1(5):206–215, 2019.

[Rus12]   B. Russell. *The Problems of Philosophy. Chap. VI. On induction*. Home Univ. Libr.; Oxford Univ.
           Pr., 1959, 1912.

[Rus15]   S. J. Russell. Unifying logic and probability. *Commun. ACM*, 58(7):88–97, 2015.

[RYM10]   S. Riedel, L. Yao, and A. McCallum. Modeling relations and their mentions without labeled
           text. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*,
           pages 148–163. Springer, 2010.

[RYMM13]   S. Riedel, L. Yao, A. McCallum, and B. M. Marlin. Relation extraction with matrix factorization and universal schemas. In *Proceedings of HLT-NAACL*, pages 74–84, 2013.

[SB19]   D. Selsam and N. Bjørner. Guiding high-performance SAT solvers with unsat-core predictions. In Janota and Lynce [JL19], pages 336–353.

[SCD18]   A. Shih, A. Choi, and A. Darwiche. A symbolic approach to explaining bayesian network classifiers. In Lang [Lan18], pages 5103–5111.

[SCMN13]   R. Socher, D. Chen, C. D. Manning, and A. Y. Ng. Reasoning with neural tensor networks for knowledge base completion. In C. J. C. Burges, L. Bottou, Z. Ghahramani, and K. Q. Weinberger, editors, *Proceedings of the 27th Annual Conference on Neural Information Processing Systems (NIPS 2013)*, pages 926–934, 2013.

[SDD+18]   S. A. Seshia, A. Desai, T. Dreossi, D. J. Fremont, S. Ghosh, E. Kim, S. Shivakumar, M. Vazquez-Chanlatte, and X. Yue. Formal specification for deep neural networks. In Lahiri and Wang [LW18], pages 20–34.

[SdG16]   L. Serafini and A. S. d'Avila Garcez. Learning and reasoning with logic tensor networks. In G. Adorni, S. Cagnoni, M. Gori, and M. Maratea, editors, *AI*IA 2016: Advances in Artificial Intelligence - Proc. XVth Int. Conf. of the Italian Association for Artificial Intelligence, Genova, Nov. 29 - Dec. 1*, volume 10037 of *LNCS*, pages 334–348. Springer, 2016.

[SDPS07]   M. Serrurier, D. Dubois, H. Prade, and T. Sudkamp. Learning fuzzy rules with their implication operators. *Data Knowl. Eng.*, 60(1):71–89, 2007.

[SFV95]   T. Schiex, H. Fargier, and G. Verfaillie. Valued constraint satisfaction problems: Hard and easy problems. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI 95)*, pages 631–639. Morgan Kaufmann, 1995.

[SGHK19]   D. Saxton, E. Grefenstette, F. Hill, and P. Kohli. Analysing mathematical reasoning abilities of neural models. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019.

[SGM+18]   G. Singh, T. Gehr, M. Mirman, M. Püschel, and M. T. Vechev. Fast and effective robustness certification. In Bengio et al. [BWL+18], pages 10825–10836.

[SGPV19a]   G. Singh, T. Gehr, M. Püschel, and M. T. Vechev. An abstract domain for certifying neural networks. *PACMPL*, 3(POPL):41:1–41:30, 2019.

[SGPV19b]   G. Singh, T. Gehr, M. Püschel, and M. T. Vechev. Boosting robustness certification of neural networks. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019.

[SGT+09]   F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini. The graph neural network model. *IEEE Trans. Neural Networks*, 20(1):61–80, 2009.

[Sha76]   G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, N.J., 1976.

[She94]   P. P. Shenoy. Conditional independence in valuation-based systems. *Int. J. Approx. Reasoning*, 10(3):203–234, 1994.

[SHH02]   H. X. S. H. Huanga. Extract intelligible and concise fuzzy rules from neural networks. *Fuzzy Sets and Systems*, 132:233–243, 2002.

[SLB+18]   D. Selsam, M. Lamm, B. Bünz, P. Liang, L. de Moura, and D. L. Dill. Learning a SAT solver from single-bit supervision. *CoRR*, abs/1802.03685, 2018.

[SLB+19]   D. Selsam, M. Lamm, B. Bünz, P. Liang, L. de Moura, and D. L. Dill. Learning a SAT solver from single-bit supervision. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019.

[SM19]   W. Samek and K. Müller. Towards explainable artificial intelligence. In Samek et al. [SMV+19], pages 5–22.

[SMV+19]   W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen, and K. Müller, editors. *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, volume 11700 of *Lecture Notes in Computer Science*. Springer, 2019.

[SN02]   J. Schumann and S. D. Nelson. Toward v&v of neural network based controllers. In D. Garlan, J. Kramer, and A. L. Wolf, editors, *Proceedings of the First Workshop on Self-Healing Systems, WOSS 2002, Charleston, South Carolina, USA, November 18-19, 2002*, pages 67–72. ACM, 2002.

[SNGR09]   R. Singh, J. P. Near, V. Ganesh, and M. Rinard. AvatarSAT: An auto-tuning boolean SAT solver. Technical Report MIT-CSAIL-TR-2009-039, MIT, 2009.

[Sow84]   J. F. Sowa. *Conceptual Structures : Information Processing in Mind and Machine*. Addison-Wesley, 1984.

[SP07]   M. Serrurier and H. Prade. Introducing possibilistic logic in ILP for dealing with exceptions. *Artif. Intell.*, 171(16-17):939–950, 2007.

[SP13a]   S. Schockaert and H. Prade. Interpolative and extrapolative reasoning in propositional theories using qualitative knowledge about conceptual spaces. *Artif. Intell.*, 202:86–131, 2013.

[SP13b]   M. Serrurier and H. Prade. An informational distance for estimating the faithfulness of a possibility distribution, viewed as a family of probability distributions, with respect to data. *Int. J. Approx. Reasoning*, 54(7):919–933, 2013.

[SP15]   M. Serrurier and H. Prade. Entropy evaluation based on confidence intervals of frequency estimates : Application to the learning of decision trees. In F. R. Bach and D. M. Blei, editors, *Proc. 32nd Int. Conf. on Machine Learning (ICML'15), Lille, July 6-11*, volume 37 of *JMLR Workshop and Conference Proceedings*, pages 1576–1584. JMLR.org, 2015.

[SQZY17]   J. Shen, Y. Qu, W. Zhang, and Y. Yu. Wasserstein distance guided representation learning for domain adaptation. *arXiv preprint arXiv:1707.01217*, 2017.

[SSBD14]   S. Shalev-Shwartz and S. Ben-David. *Understanding Machine Learning. From theory to algorithms*. Cambridge University Press, 2014.

[SSZ+17]   G. Sourek, M. Svatos, F. Zelezný, S. Schockaert, and O. Kuzelka. Stacked structure learning for lifted relational neural networks. In *Inductive Logic Programming - 27th International Conference, ILP 2017, Orléans, France, September 4-6, 2017, Revised Selected Papers*, pages 140–151, 2017.

[SV08]   G. Shafer and V. Vovk. A tutorial on conformal prediction. *Journal of Machine Learning Research*, 9(Mar):371–421, 2008.

[SY05]   N. Stroppa and F. Yvon. Analogical learning and formal proportions: Definitions and methodological issues. Technical Report D004, ENST-Paris, 2005.

[SZS+14]   C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. J. Goodfellow, and R. Fergus. Intriguing properties of neural networks. In Y. Bengio and Y. LeCun, editors, *2nd International Conference on Learning Representations, ICLR 2014, Banff, April 14-16, 2014, Conf. Track Proc.*, 2014.

[TCP+15] K. Toutanova, D. Chen, P. Pantel, H. Poon, P. Choudhury, and M. Gamon. Representing text for joint embedding of text and knowledge bases. In *Proceedings of EMNLP*, pages 1499–1509, 2015.

[Tha78] P. R. Thagard. The best explanation: Criteria for theory choice. *The Journal of Philosophy*, 75(2):76–92, 1978.

[Tib96] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

[TMLZ18] G. Tao, S. Ma, Y. Liu, and X. Zhang. Attacks meet interpretability: Attribute-steered detection of adversarial samples. In Bengio et al. [BWL+18], pages 7728–7739.

[TS85] T. Takagi and M. Sugeno. Fuzzy identification of systems and its application to modelling and control. *IEEE Trans. Systems, Man, and Cybernetics*, 15(1):11)–132, 1985.

[TWR+16] T. Trouillon, J. Welbl, S. Riedel, É. Gaussier, and G. Bouchard. Complex embeddings for simple link prediction. In *Int. Conf. on Machine Learning*, pages 2071–2080, 2016.

[TWX+18] R. Tomsett, A. Widdicombe, T. Xing, S. Chakraborty, S. Julier, P. Gurram, R. M. Rao, and M. B. Srivastava. Why the failure? how adversarial examples can provide insights for interpretable machine learning. In *21st International Conference on Information Fusion, FUSION 2018, Cambridge, UK, July 10-13, 2018*, pages 838–845. IEEE, 2018.

[UVS11] J. Urban, J. Vyskocil, and P. Stepánek. Malecop machine learning connection prover. In K. Brünnler and G. Metcalfe, editors, *Automated Reasoning with Analytic Tableaux and Related Methods - 20th International Conference, TABLEAUX 2011, Bern, Switzerland, July 4-8, 2011. Proceedings*, volume 6793 of *LNCS*, pages 263–277. Springer, 2011.

[Val84] L. G. Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.

[Vap13] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer, 2013.

[VC71] V. N. Vapnik and A. Y. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and Its Applications*, 16(2):264–280, 1971.

[VFJ15] O. Vinyals, M. Fortunato, and N. Jaitly. Pointer networks. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28: Annual Conf. on Neural Information Processing Systems, Dec. 7-12, Montreal*, pages 2692–2700, 2015.

[VM15] L. Vilnis and A. McCallum. Word representations via gaussian embedding. In *Proceedings of the International Conference on Learning Representations*, 2015.

[VNP+19] H. Verhaeghe, S. Nijssen, G. Pesant, C.-G. Quimper, and P. Schaus. Learning optimal decision trees using constraint programming. *Constraints*, 2019.

[VZ17] S. Verwer and Y. Zhang. Learning decision trees with flexible constraints and objectives using integer optimization. In D. Salvagnin and M. Lombardi, editors, *Integration of AI and OR Techniques in Constraint Programming - Proc. 14th Int. Conf. CPAIOR'17, Padua, Italy, June 5-8*, volume 10335 of *LNCS*, pages 94–103. Springer, 2017.

[VZ19] S. Verwer and Y. Zhang. Learning optimal classification trees using a binary linear program formulation. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019* [AAA19], pages 1625–1632.

[WBD+19] C. Wang, R. Bunel, K. Dvijotham, P. Huang, E. Grefenstette, and P. Kohli. Knowing when to stop: Evaluation and verification of conformity to output-size specifications. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pages 12260–12269. Computer Vision Foundation / IEEE, 2019.

[WCN96] W. Wen, J. Callahan, and M. Napolitano. Towards developing verifiable neural network controller. In *Workshop on AI for Aeronautics and Space*, 1996.

[WDWK19] P. Wang, P. L. Donti, B. Wilder, and J. Z. Kolter. SATNet: bridging deep learning and logical reasoning using a differentiable satisfiability solver. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, pages 6545–6554, 2019.

[Wha16] D. Whalen. Holophrasm: a neural automated theorem prover for higher-order logic. *CoRR*, abs/1608.02644, 2016.

[Wil06] N. Wilson. An efficient upper approximation for conditional preference. In G. Brewka, S. Coradeschi, A. Perini, and P. Traverso, editors, *Proceedings of the 17th European Conference on Artificial Intelligence (ECAI 2006)*, Frontiers in Artificial Intelligence and Applications. IOS Press, 2006.

[Wil09] N. Wilson. Efficient inference for expressive comparative preference language. In C. Boutilier, editor, *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI'09)*, pages 961–966, 2009.

[Wil14] N. Wilson. Preference inference based on lexicographic models. In T. Schaub, G. Friedrich, and B. O'Sullivan, editors, *Proceedings of th 21st European Conference on Artificial Intelligence (ECAI 2014)*, volume 263 of *Frontiers in Artificial Intelligence and Applications*, pages 921–926. IOS Press, 2014.

[WRTK19] S. Webb, T. Rainforth, Y. W. Teh, and M. P. Kumar. A statistical approach to assessing neural network robustness. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, May 6-9, 2019*. OpenReview.net, 2019.

[WTWD17] M. Wang, Y. Tang, J. Wang, and J. Deng. Premise selection for theorem proving by deep graph embedding. In Guyon et al. [GvLB+17], pages 2786–2796.

[XDZD16] P. Xu, F. Davoine, H. Zha, and T. Denœux. Evidential calibration of binary SVM classifiers. *Int. J. of Approximate Reasoning*, 72:55–70, 2016.

[XHHL08] L. Xu, F. Hutter, H. H. Hoos, and K. Leyton-Brown. SATzilla: Portfolio-based algorithm selection for SAT. *J. Artif. Intell. Res.*, 32:565–606, 2008.

[XHMZ17] H. Xiao, M. Huang, L. Meng, and X. Zhu. Ssp: Semantic space projection for knowledge graph embedding with text descriptions. In *Proceedings of AAAI*, volume 17, pages 3104–3110, 2017.

[XKK18] H. Xu, S. Koenig, and T. K. S. Kumar. Towards effective deep learning for constraint satisfaction problems. In Hooker [Hoo18], pages 588–597.

[XKS92] L. Xu, A. Krzyzak, and C. Y. Suen. Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Trans. on Systems, Man and Cybernetics*, 22(3):418–435, 1992.

[XLJ+16] R. Xie, Z. Liu, J. Jia, H. Luan, and M. Sun. Representation learning of knowledge graphs with entity descriptions. In *Proceedigs of AAAI*, pages 2659–2665, 2016.

[XLZ+18] K. Xu, S. Liu, P. Zhao, P. Chen, H. Zhang, D. Erdogmus, Y. Wang, and X. Lin. Structured adversarial attack: Towards general implementation and better interpretability. *CoRR*, abs/1808.01664, 2018.

[XUD+19]   F. Xu, H. Uszkoreit, Y. Du, W. Fan, D. Zhao, and J. Zhu. Explainable AI: A brief survey on history, research areas, approaches and challenges. In *NLPCC*, pages 563–574, 2019.

[XXM+19]   Y. Xie, Z. Xu, K. Meel, M. S. Kankanhalli, and H. Soh. Semantically-regularized logic graph embeddings. *CoRR*, abs/1909.01161, 2019.

[XZF+18]   J. Xu, Z. Zhang, T. Friedman, Y. Liang, and G. V. den Broeck. A semantic loss function for deep learning with symbolic knowledge. In *Proc. 35th Int. Conf. on Machine Learning*, pages 5498–5507, 2018.

[YD19]   K. Yang and J. Deng. Learning to prove theorems via interacting with proof assistants. In K. Chaudhuri and R. Salakhutdinov, editors, *Proc 36th Int. Conf. on Machine Learning (ICML'19), 9-15 June, Long Beach, California*, volume 97 of *Proc. of Machine Learning Research*, pages 6984–6994. PMLR, 2019.

[YLX+18]   K. Yoon, R. Liao, Y. Xiong, L. Zhang, E. Fetaya, R. Urtasun, R. S. Zemel, and X. Pitkow. Inference in probabilistic graphical models by graph neural networks. In *6th Int. Conf. on Learning Representations (ICLR'18), Vancouver, April 30 - May 3, Workshop Track Proc.*, 2018.

[YM08]   B. B. Yaghlane and K. Mellouli. Inference in directed evidential networks based on the transferable belief model. *Int. J. Approx. Reasoning*, 48(2):399–418, 2008.

[YWC+19]   Z. Yang, F. Wang, Z. Chen, G. Wei, and T. Rompf. Graph neural reasoning for 2-quantified boolean formula solvers. *CoRR*, abs/1904.12084, 2019.

[YYC17]   F. Yang, Z. Yang, and W. W. Cohen. Differentiable learning of logical rules for knowledge base reasoning. In Guyon et al. [GvLB+17], pages 2319–2328.

[Zad73]   L. A. Zadeh. Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Trans. Systems, Man, and Cybernetics*, 3(1):28–44, 1973.

[Zad79]   L. A. Zadeh. A theory of approximate reasoning. In J. E. Hayes, D. Mitchie, and L. L. Mikulich, editors, *Machine intelligence, Vol. 9*, pages 149–194. Ellis Horwood, 1979.

[Zad92]   L. A. Zadeh. The calculus of fuzzy if-then rules. *AI Expert*, 7(3):22–27, 1992.

[Zak01]   R. R. Zakrzewski. Verification of a trained neural network accuracy. In *International Joint Conference on Neural Networks (IJCNN)*, pages 1657–1662. IEEE, 2001.

[ZBD19]   H. Zhang, S. Basu, and I. Davidson. Deep constrained clustering - algorithms and advances. *CoRR*, abs/1901.10061, 2019.

[ZD98]   L. M. Zouhal and T. Denœux. An evidence-theoretic k-NN rule with parameter optimization. *IEEE Trans. on Systems, Man and Cybernetics C*, 28(2):263–271, 1998.

[Zho19]   Z. Zhou. Abductive learning: towards bridging machine learning and logical reasoning. *SCIENCE CHINA Information Sciences*, 62(7):76101:1–76101:3, 2019.

[ZKL+15]   B. Zhou, A. Khosla, À. Lapedriza, A. Oliva, and A. Torralba. Learning deep features for discriminative localization. *CoRR*, abs/1512.04150, 2015.

[ZL06]   F. Zelezný and N. Lavrac. Propositionalization-based relational subgroup discovery with rsd. *Machine Learning*, 62(1-2):33–63, 2006.

[ZRF+18a]   L. Zhang, G. Rosenblatt, E. Fetaya, R. Liao, W. E. Byrd, M. Might, R. Urtasun, and R. S. Zemel. Neural guided constraint logic programming for program synthesis. In Bengio et al. [BWL+18], pages 1744–1753.

[ZRF⁺18b] L. Zhang, G. Rosenblatt, E. Fetaya, R. Liao, W. E. Byrd, R. Urtasun, and R. S. Zemel. Leveraging constraint logic programming for neural guided program synthesis. In *6th Int. Conf. on Learning Representations (ICLR'18), Vancouver, April 30 - May 3, Workshop Track Proc.* OpenReview.net, 2018.

[ZXMJ19] H. Zhu, Z. Xiong, S. Magill, and S. Jagannathan. An inductive synthesis framework for verifiable reinforcement learning. In K. S. McKinley and K. Fisher, editors, *Proceedings of the 40th ACM SIGPLAN Conference on Programming Language Design and Implementation, PLDI 2019, Phoenix, AZ, USA, June 22-26, 2019*, pages 686–701. ACM, 2019.

[ZZW⁺15] H. Zhong, J. Zhang, Z. Wang, H. Wan, and Z. Chen. Aligning knowledge and text embeddings by entity descriptions. In *Proceedings of EMNLP*, pages 267–272, 2015.