SCIENCE CHINA
Information Sciences

• PERSPECTIVE •

July 2019, Vol. 62 076101:1–076101:3
https://doi.org/10.1007/s11432-018-9801-4

Abductive learning: towards bridging machine learning and logical reasoning

Zhi-Hua ZHOU

National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023*, China*

Received 21 December 2018/Accepted 15 February 2019/Published online 8 March 2019

Citation Zhou Z-H. Abductive learning: towards bridging machine learning and logical reasoning. Sci China Inf Sci, 2019, 62(7): 076101, https://doi.org/10.1007/s11432-018-9801-4

In the history of artificial intelligence research, machine learning and logical reasoning have almost been separately developed. It is often argued that advanced intelligent technologies would emerge when machine learning and logical reasoning are seamlessly integrated as human beings generally perform problem-solving based on the leverage of perception and reasoning, where perception corresponds to a data-driven process that can be realized by machine learning whereas reasoning corresponds to a knowledge-driven process that can be realized by logical reasoning. Developing a unified framework which accommodates and enables machine learning and logical reasoning to work together has been deemed as the holy grail challenge for the artificial intelligence community.

The major difficulty lies in the fact that popular reasoning techniques are generally based on first-order logic representation, but popular learning techniques do not. Efforts for bridging machine learning and logical reasoning generally try to adapt one to the other. For example, probabilistic logic program (PLP) [1] attempts to extend first-order logic to accommodate probabilistic groundings such that probabilistic inference can be included, whereas statistical relational learning (SRL) [2] attempts to construct/initialize a probabilistic model based on domain knowledge expressed in first-order logic clauses. PLP follows a "heavy-reasoning light-learning" way that preserves the strength of logical reasoning but does not fully exploit machine learning, whereas SRL adopts a "heavy-learning light-reasoning" way that preserves the strength of machine learning but does not fully exploit logical reasoning.

In this study, we propose abductive learning, a new framework towards bridging machine learning and logical reasoning. The abduction, also known as retro-production, refers to the process of selectively inferring certain facts and hypotheses that explain phenomena and observations based on background knowledge [3]. Given observed facts and background knowledge expressed as first-order logic clauses, logical formalization of abductive reasoning, i.e., logical abduction, can abduce hypotheses as possible explanations to the observed facts. Abduction has been a recurring topic of interest in the field of artificial intelligence, and in the history it had been tried to integrate with symbolic induction [4, 5]. Recently it has been shown beneficial to exploit logical abduction to introduce symbolic knowledge into statistical learning [6]. Considering that modern machine learning is also a kind of induction, it is interesting to leverage induction and abduction in a mutually beneficial way in a unified framework.

In conventional supervised learning, e.g., classification, we are given a set of training examples $\{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_m, y_m)\}$ where $\boldsymbol{x}_i \in \mathcal{X}$ is the i-th training instance, $y_i \in \mathcal{Y}$ is the ground-truth class label, and the task is to learn a function $f : \mathcal{X} \mapsto \mathcal{Y}$ from the input space $\mathcal{X}$ to the output space $\mathcal{Y}$ to accurately predict the unseen data. The setting of abductive learning is a bit differ-

Email: zhouzh@nju.edu.cn

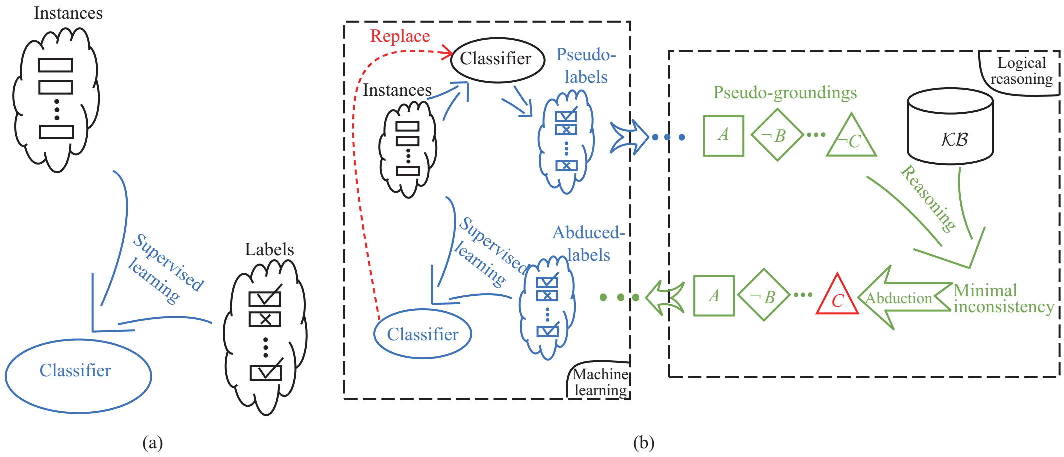

Figure 1 (a) Conventional supervised learning where the ground-truth labels of training data are given and (b) abductive learning where a classifier and a knowledge base are given. The given information is highlighted in black; the machine learning and logical reasoning components are shown in blue and green, respectively. In (b), the given classifier helps generate pseudo-labels, leading to pseudo-groundings; then, revisions to the pseudo groundings (shown as a red triangle) are generated via logical abduction based on minimizing the inconsistency with the knowledge base. The abduced labels are used to train the classifier, which is then adopted to replace the original classifier in the next iteration.

ent: We are given $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m\}$, a knowledge base $\mathcal{KB}$, and a classifier $\mathcal{C}$; the task is to learn a function f for making correct predictions on unseen data, and meanwhile, the logical facts grounded by $\{(\boldsymbol{x}_1, f(\boldsymbol{x}_1)), \ldots, (\boldsymbol{x}_m, f(\boldsymbol{x}_m))\}$ should be compatible with $\mathcal{KB}$. Formally, given $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_i\}$, $\mathcal{C}$ and $\mathcal{KB}$, the task is to seek a function f such that

$$\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_i\},\ f \rhd \mathcal{O} \tag{1}$$

$$\text{s.t.} \quad \mathcal{KB} \models \mathcal{O}, \quad \text{or} \tag{2}$$

$$\mathcal{KB} \models \Delta(\mathcal{O}), \quad f \leftarrow \psi(f, \Delta(\mathcal{O})), \tag{3}$$

where $\mathcal{O}$ denotes the logical facts grounded by $\boldsymbol{x}_i$'s and f, and $\models$ denotes logical entailment. If $\mathcal{O}$ is consistent with $\mathcal{KB}$ as shown in (2), the current f will be returned; otherwise, $\Delta(\mathcal{O})$ will be generated via logical abduction as specified by the first equation in (3), and f will then be updated based on $\Delta(\mathcal{O})$ by a process ψ. If $\mathcal{O}$ is not compatible with $\mathcal{KB}$, i.e., $\mathcal{KB} \not\models \mathcal{O}$ and there does not exist Δ enabling $\mathcal{KB} \models \Delta(\mathcal{O})$, the process terminates and return `False`.

Figure 1 provides a simple illustration. In contrast to conventional supervised learning, the proposed abductive learning does not rely on ground-truth labels (although the existence of ground-truth labels will make it easier as will be discussed later). Instead, the learning is facilitated with a classifier and a knowledge base containing first-order logic clauses. Predictions made by the classifier are used as pseudo-labels of the training instances, leading to pseudo-grounded facts such as the $A, \neg B, \ldots, \neg C$ shown in Figure 1(b), corresponding to $\mathcal{O}$ in (1). Then, a logical reasoning process is applied to verify whether these pseudo-groundings are consistent with the knowledge base. If the findings are consistent, corresponding to the situation implied by (2), the learning process terminates and the current classifier, f, is returned. Otherwise, a hypothetical revision to the pseudo-groundings, based on minimizing the inconsistency with the knowledge base, will be generated via logical abduction. For example, suppose that the $\neg C$ in the pseudo-groundings in Figure 1(b) has been hypothetically suggested to change to C as the minimal revision, corresponding to $\Delta(\mathcal{O})$ in (3), to make the groundings consistent with the knowledge base. Then, the revision is performed, leading to the modification of some pseudo-labels to generate the abduced-labels. The abduced-labels will be used like ground-truth labels in conventional supervised learning to train the classifier which will then be adopted to replace the original classifier in the next iteration, corresponding to the ψ process in (3).

Some remarks are to be made. First, although Eq. (1) and Figure 1(b) are not based on the assumption of the existence of ground-truth labels, this does not imply that abductive learning cannot utilize ground-truth labels. Actually, if some ground-truth labels are available, the pseudo-grounded facts are expected to be more reliable and the logical abduction can be more effective because the ground-truth labels help prune the hypothesis space for abduction. Ground-truth labels can also be used to train the final classifier, leading to potential improvement of performance. Indeed, abductive learning can be regarded as a special kind of weakly supervised learning [7], where the supervision information comes from knowledge reasoning. Evidently, supervision can be enhanced if ground-truth labels are given in addition to the knowledge base.

Second, the initial classifier can be simple, such as preprocessing based on clusterings or nearest neighbor classification. In practice, it is more amenable to be a pre-trained model, e.g., a deep learning model pre-trained from benchmark datasets. It can also be a result of transfer learning from relevant tasks, or even a reusable learnware [8]. If abundant labeled data are given, the initial classifier is not required because the logical groundings can be directly generated by the labeled data; in this sense, labeled data can be regarded as an alternative to an initial classifier.

Third, Eqs. (2) and (3) and Figure 1(b) assume that the given knowledge base is accurate and consistent; thus, only revisions of pseudo-groundings will be considered to minimize the inconsistency with the knowledge base. In real practice, however, the given knowledge base may be inaccurate or may contain contradicting clauses; therefore, the knowledge base itself is subject to refinement. This will make the abductive learning process more complicated, and many technical challenges should be addressed. Actually, even if the knowledge base is assumed to be accurate and consistent, minimizing the inconsistency remains a big challenge because it involves optimization on symbolic relations rather than the commonly used numerical optimization.

The conversion of pseudo-labels to pseudo-groundings, and vice versa, is dependent on the realization. For example, in the neural logical machine (NLM) [9], an implementation of abductive learning, the conversion was simply made by regarding the labels predicted by the classifier as arguments in predicates of first-order logic clauses; therein, the classifier and logical reasoning were realized by a convolutional neural network and abductive logic programming, respectively. The machine learning component is realized by convolutional neural networks; therefore, the resulting NLM model is related to neural-symbolic learning systems [10]. However, the abductive learning framework is highly general and flexible and other machine learning mechanisms, besides deep learning, can be employed. The logical abduction can be similarly realized via other logical reasoning mechanisms and need not be limited to abductive logic programming.

From a psychological viewpoint, the perception and reasoning of human beings are entangled rather than separated; this is reassembled to some sense in abductive learning. Indeed, human perception is not always accurate, and some misperceptions can be corrected when the perceived patterns are involved in reasoning, wherein the reasoning process can provide a "guess" about the correct patterns. Such guesses are simulated by the abducted-labels in Figure 1(b). Human knowledge can be updated/refined when new facts are observed; this can be simulated via (potential) knowledge refinement in abductive learning.

In summary, abductive learning provides a new framework, in which machine learning and logical reasoning can be entangled and mutually beneficial. This framework is quite general, with numerous avenues for future investigations. The exploration of abductive learning will possibly provide new approaches for developing a unified framework that accommodates learning and reasoning.

Acknowledgements This work was supported by National Key R&D Program of China (Grant No. 2018YFB1004300) and National Natural Science Foundation of China (Grant No. 61751306).

References

1 de Raedt L, Frasconi P, Kersting K, et al. Probabilistic Inductive Logic Programming. Berlin: Springer, 2008
2 Getoor L, Taskar B, eds. Introduction to Statistical Relational Learning. Cambridge: MIT Press, 2007
3 Magnani L. Abductive Cognition: the Epistemological and Eco-Cognitive Dimensions of Hypothetical Reasoning. Berlin: Springer, 2009
4 Mooney R J. Integrating abduction and induction in machine learning. In: Abduction and Induction. Dordrecht: Kluwer Academic Publishers, 2000. 181–191
5 Muggleton S H, Bryant C H. Theory completion using inverse entailment. In: Proceedings of the 10th International Conference on Inductive Logic Programming, London, 2000. 130–146
6 Dai W-Z, Zhou Z-H. Combining logic abduction and statistical induction: discovering written primitives with human knowledge. In: Proceedings of the 31st AAAI Conference on Artificial Intelligence, San Francisco, 2017. 2977–2983
7 Zhou Z-H. A brief introduction to weakly supervised learning. Natl Sci Rev, 2018, 5: 44–53
8 Zhou Z-H. Learnware: on the future of machine learning. Front Comput Sci, 2016, 10: 589–590
9 Dai W-Z, Xu Q-L, Yu Y, et al. Tunneling neural perception and logic reasoning through abductive learning. 2018. ArXiv: 1802.01173
10 Garcez A S D, Broda K B, Gabbay D M. Neural-Symbolic Learning Systems: Foundations and Applications. London: Springer, 2009