Logics for Neural Network Learning

Caleb Kisby

Dissertation Proposal for the Degree of Doctor of Philosophy
Department of Computer Science
Indiana University

Abstract. Artificial intelligence has long been divided by two of its major paradigms: connectionist learning and symbolic reasoning. Almost all neuro-symbolic proposals treat learning as a black-box process that occurs before, after, or within a symbolic reasoner; as of yet, how learning could seamlessly interface reasoning remains a mystery. I propose a neuro-symbolic interface that bridges neural network learning with symbolic reasoning. In particular, I claim that learning policies that neural networks use in practice (specifically Hebbian learning and backpropagation) formally correspond to dynamic operators in human-readable modal logics. I provide a plan for proving this correspondence via a series of soundness and completeness theorems. Moreover, these theorems give rise to neuro-symbolic translation algorithms, which I intend to implement using Tensorflow.

Introduction

Humans make intelligent decisions by seamlessly integrating both their ability to learn and their ability to reason about what they have learned. But researchers in artificial intelligence have long experienced a tradeoff between the two: Neural systems have had tremendous success learning from unstructured data, whereas symbolic systems excel at sophisticated reasoning tasks that neural systems cannot readily learn. In the last three decades, there have been countless hybrid systems that combine neural and symbolic components in a myriad of ways, hoping to strike the right balance [1, 6, 15, 22, 25]. Other authors [2, 5, 8, 7, 10, 16, 17] suggest a more principled approach: The neural and symbolic are two ways of interpreting the same agent, and we should be able to translate between the two. In fact, Garcez et. al. [8, 7] has demonstrated that we can extract (sound) knowledge from a net, as well as build a net that (completely) models existing knowledge. Equivalently, Leitgeb [16, 17] viewed neural networks as the semantics of a formal logic, and showed that this logic completely axiomatizes the behavior of the net.

However, no existing neuro-symbolic proposal seamlessly interfaces learning and reasoning like humans do. Most neuro-symbolic hybrid systems, including that of Garcez et. al., treat learning as a black-box process that occurs before, after, or within a symbolic reasoner. In addition, more formal translations such as Leitgeb's do not even consider learning. As of yet, how learning relates to reasoning remains a mystery. This leads us to the central claim of my dissertation:

Thesis: The learning that neural networks use in practice
formally corresponds to a dynamic operator in a particular modal logic.

What I have in mind is a modal logic with (at least) a dynamic operator $[\varphi]\psi$ whose semantics are given by "after learning φ, evaluate ψ." Such a logic serves as a bridge between neural network learning and formal reasoning: Its *semantics* are given by learning dynamics in a neural network, whereas its *rules of inference* encode formal reasoning about what has been learned.

By "formally corresponds," I mean that we are after the soundness and completeness of this logic. That is, a neural network models a sentence φ if and only if φ follows from the rules of inference. I hypothesize that we can write down a complete axiomatization of neural network learning using only well-understood tools from modal logic taken from off-the-shelf. That is not to say that the modal logic needs to be standard (normal, monotonic), but its inferences must be expressed in an intuitive (human-readable) language.

My plan is to demonstrate this in a simplified setting, with binary networks using Hebbian learning as a proof-of-concept. But an important part of my dissertation will be to generalize this logic of Hebbian learning to account for neural networks used in practice. In particular, I hypothesize that we can tell a similar story for neural networks with fuzzy (real) activation values that learn via backpropagation.

An important follow-up question is what we gain from this correspondence. First, we should point out that it is already possible, via [8, 7], to take some prior knowledge Γ, construct a feed-forward neural network $\mathcal{N}$ that models Γ, train it to obtain $\mathcal{N}'$, and then extract its knowledge Γ'. This is a landmark acheivement that goes a long way towards helping us understand a neural network's knowledge post-training. Completeness of our modal logic similarly gives us neural network extraction and model building, but with a twist: We can inject statements of the form $[\varphi]\psi$ into Γ, specifying what we would like the net to learn! As an integral part of my dissertation, I will write open-source code (using Tensorflow [18]) that implements this construction.

Finally, and from a more philosophical point of view, this correspondence teaches us about the nature of neural network learning. Specifically, I intend to show that we should understand Hebbian learning and backpropagation as a kind of *preference upgrade*, in the sense of [23, 24]. Moreover, learning changes the network's preferences according to our logic's rules of inference.

Related Work

Neuro-Symbolic AI

Until recently, artificial intelligence has been divided by two of its major paradigms: connectionist learning and symbolic reasoning. But the last three decades have seen a monumental effort to achieve the best of both worlds. The field of *neuro-symbolic artificial intelligence* aims to integrate the two in a way that preserves both the flexibility of learning and the strength of reasoning. The space of neuro-symbolic systems is vast, so I will focus here on a few key trends. (See [1] for a survey at the field's inception, and [22, 25] for more recent surveys.)

In his famous AAAI 2020 Robert S. Engelmore Memorial Award Lecture on the state of artificial intelligence, Henry Kautz offers six possible architectures for the design of a neuro-symbolic system [13]. Although not comprehensive, this list captures the most common hybrid architectures to date. In turn:

Symbolic Neuro Symbolic. Symbolic input is converted into vectors, fed into a trained neural network, and then the resulting prediction is converted to symbolic output.

Symbolic[Neuro]. A trained neural subsystem operates within a symbolic system.

Neuro | Symbolic. A trained neural system assists a symbolic system, usually by converting input to symbolic form.

Neuro: Symbolic → Neuro. Same as Symbolic Neuro Symbolic architecture, but the neural network is specially trained with the symbolic rules in mind.

Neuro$_{\textbf{Symbolic}}$. A neural network is composed of submodules that are constructed based on symbolic rules.

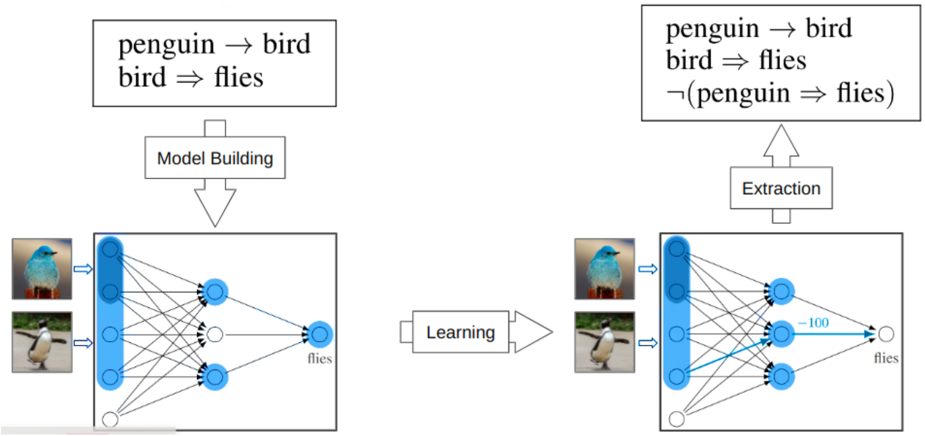

Figure 1. Given an incomplete conditional knowledge base, we can construct an equivalent feed-forward neural network. In this example, the network believes that all birds, including penguins, fly. We then train the net, and see (via extraction) that it now believes that penguins are an exception to the rule.

Neuro[Symbolic]. A symbolic subsystem operates within a neural system.

Although these architectures combine neural networks with symbolic reasoners, notice — crucially — that none of them *interface* learning with reasoning. In fact, with the exception of Neuro[Symbolic], all of these architectures involve learning as a black-box process that occurs before, after, or within a symbolic reasoner. As for the Neuro[Symbolic] architecture, Kautz does not consider how the training of the neural network might affect the internal symbolic reasoner. In general, neuro-symbolic AI has offered few ideas to illuminate the relationship between learning and reasoning; how we should interface the two remains a mystery.

Formal Correspondences Between Nets and Logic

Not mentioned in Kautz' lecture is a more principled, foundational approach to bridging the neural and symbolic. The key idea is that rather than trying to combine specific neural and symbolic systems, we can instead view the neural and symbolic as two perspectives we can have about the same agent. In other words, the two are just different representations of the same mental content.

The most well-known proponents of this view are perhaps Garcez et. al. Their contribution comes in two parts. First, they provide an algorithm to extract rules from a neural network (*neural network extraction*) [7]. These rules are expressed using $\wedge, \vee$, and $\neg$, as well as negation by default $\sim\varphi$ ("not φ holds, unless we can derive φ"), which the authors argue is necessary to extract any meaningful rules from a neural network. For feed-forward networks, this algorithm is provably sound and complete: The neural network models a sentence φ if and only if φ follows from the extracted rules.

Their second landmark achievement [8] is an algorithm for doing the inverse. Given a set of rules Γ, this algorithm constructs a neural network that models Γ (*model building*). They provide an algorithm for each of several non-classical logics (including modal, temporal, and intuitionistic logics). But the reasoning system underlying all of these is the propositional logic of $\wedge, \vee, \neg$, with negation by default $\sim\varphi$.

Put together, these two algorithms provide a way to translate between neural and symbolic representations. This allows us to constrain neural networks by specifying constraints they should obey (at least prior to training). Additionally, we can read off the beliefs of neural networks post-training. As shown in Figure 1: Given a set of prior knowledge Γ, we can construct a feed-forward neural network $\mathcal{N}$ that models Γ, train it to obtain $\mathcal{N}'$, and then extract its knowledge Γ'.

A less well-studied, yet equivalent, approach is to view the behavior of neural networks as the semantics for a symbolic logic. In fact, this idea dates back to McCulloch and Pitts' seminal paper [19]. But this approach has had a modern reimagining, á la the work of Balkenius, Gärdenfors [2], and later Leitgeb [16, 17] (and recently has been independently discovered by Giordano et. al. [10]). The basic idea is to first assign a set of neurons S to each base proposition p. The meaning of more complex formulas is determined compositionally as follows:

$\varphi \wedge \psi$	is	the union of neurons for φ and ψ
$\varphi \vee \psi$	is	the intersection of neurons for φ and ψ
$\varphi \rightarrow \psi$	is true iff	φ contains ψ
$\varphi \Rightarrow \psi$	is true iff	the forward propagation of φ contains ψ

Here, $\varphi \rightarrow \psi$ is material implication (read "φ implies ψ"), and $\varphi \Rightarrow \psi$ is a default conditional (read "typically, if φ then ψ"). Moreover, Leitgeb [16] showed that this neural interpretation $\varphi \Rightarrow \psi$ (in a feed-forward net) is sound and complete with respect to the rules of loop-cumulative conditionals:

$$\begin{array}{llll}
\textbf{(LLE)} & \dfrac{\vdash \varphi \leftrightarrow \psi \qquad \varphi \Rightarrow \rho}{\psi \Rightarrow \rho} & \textbf{(CCut)} & \dfrac{\varphi \wedge \psi \Rightarrow \rho \qquad \varphi \Rightarrow \psi}{\varphi \Rightarrow \rho} \\
\textbf{(RW)} & \dfrac{\vdash \varphi \rightarrow \psi \qquad \rho \Rightarrow \varphi}{\rho \Rightarrow \psi} & \textbf{(CMono)} & \dfrac{\varphi \Rightarrow \psi \qquad \varphi \Rightarrow \rho}{\varphi \wedge \psi \Rightarrow \rho} \\
\textbf{(Loop)} & \dfrac{\varphi_0 \Rightarrow \varphi_1, \varphi_1 \Rightarrow \varphi_2, \ldots, \varphi_n \Rightarrow \varphi_0}{\varphi_0 \Rightarrow \varphi_n} & &
\end{array}$$

That is, a neural network's propagation of a signal φ contains ψ if and only if $\varphi \Rightarrow \psi$ follows from these conditional rules. This makes precise the following insight:

Forward propagation in a feed-forward net
formally corresponds to a default conditional.

This is essentially what Garcez et. al. had discovered, albeit with default conditionals rather than negation by default. And as with the work of Garcez et. al., Leitgeb's proof of completeness gives us a method to construct a neural network $\mathcal{N}$ modelling a set Γ. (Similarly, soundness gives us neural network extraction.)

This thread of research suggests a general method for relating neural network dynamics with logical operators. And so it seems promising for the problem of interfacing neural learning with symbolic reasoning. However, both Garcez et. al.'s translation algorithms and the conditional logic of forward propagation lack a logical account of learning. The issue appears to be that the *language* of these logics are necessarily static; to express learning *within* the logic itself, we need new ideas.

Dynamic Epistemic Logic to the Rescue

There has been a recent turn in formal logic towards dynamic logics, i.e. logics with syntactic operators that express change of information. In the past two decades, there has been a significant effort (especially at the Institute for Logic, Language and Computation (ILLC) at the University of Amsterdam) to integrate this idea into epistemic multi-modal logics. The resulting Dynamic Epistemic Logics (DELs) model agents that change their *knowledge*, *beliefs*, and *preferences* [9].

A prototypical DEL is the logic of public announcement [21], which incorporates a dynamic modality $[\varphi!]$ ("announce φ as factually true") alongside modalities for group knowledge. This is completely hard information change — in an impossible worlds model, we eliminate all $\neg\varphi$-worlds from consideration. But DELs can model soft information change as well. For example, the modality $[\Uparrow\varphi]$ (see [23]) represents lexicographic preference upgrade, i.e. we prefer φ-worlds over $\neg\varphi$-worlds but otherwise leave the order intact.

Recent work by Baltag et. al. [3, 4] apply this DEL approach to learning. The former presents a dynamic multimodal logic that captures an agent's learning in the limit, whereas the latter models supervised learning as a game played between student and teacher. This appears to be exactly what we need: Logics that express learning in the language itself. But unfortunately, it is unclear at this stage how these logics relate to the learning policies used for neural networks (e.g. backpropagation).

Plan of Attack

The central goal of my dissertation is to provide a neuro-symbolic interface that bridges the learning neural networks use in practice with symbolic reasoning. My vision is that we can marry the methodology of DEL to the neural-correspondence approach described above. I claim that learning policies for neural networks used in practice formally correspond to DEL-style modalities $[\varphi]\psi$ ("after learning φ, evaluate ψ"). Moreover, I hypothesize that we can get a complete axiomatization of $[\varphi]\psi$ using well-known tools from modal logic.

My plan of attack is as follows. In order to apply the DEL methodology, I must first take the conditional logic of forward-propagation in a feed-forward net and convert it to a *modal* logic. I will then prove soundness and completeness for this base modal logic. This is a technical point, but the idea is that rather than focusing on statements $\varphi \Rightarrow \psi$ ("typically, if φ then ψ"), we would like to instead make $\mathbf{T}\varphi$ ("typically φ") the primitive operation. (Note that $\varphi \Rightarrow \psi$ is just $\mathbf{T}\varphi \rightarrow \psi$.) The point is that this modal language has native support for dynamic modal operators $[\varphi]\psi$.

Next, I will illustrate my thesis in a simplified setting, using binary neural networks that learn via the simplest policy we know: Naïve (unstable) Hebbian learning. I will introduce the operator $[\varphi]_{\text{hebb}}$ into the logic, and try to prove soundness and completeness. I will then generalize the story for $[\varphi]_{\text{hebb}}$ to neural networks with fuzzy (real) activation values. Finally, I will tackle the formalization of backpropagation $[\varphi;\psi]_{\text{bp}}$ using my work on Hebbian learning as a guide.

Step 1. Convert the conditional logic into a modal logic

We restate the key insight of Balkenius, Gärdenfors [2], and Leitgeb [16, 17]: Forward propagation in a feed-forward net formally corresponds to a default conditional. I would like to preserve this insight in our logic, but the language of conditionals is inadequate for my purposes. The reason is that integrating conditionals alongside belief revision is a controversial and open issue [11]. Instead, I would like to use a modal language, which will allow us to natively employ dynamic operators in the DEL fashion.

Fortunately, this just involves a straightforward semantic transformation. We can assign sets of neurons S (in a feed-forward net) to propositions p as before, and then build up formulas as follows:

$\varphi \wedge \psi$	is	the union of neurons for φ and ψ
$\varphi \vee \psi$	is	the intersection of neurons for φ and ψ
$\varphi \rightarrow \psi$	is	$\neg\varphi \vee \psi$
$\mathbf{T}\varphi$	is	the forward propagation of φ

Step 2. Prove soundness and completeness for the base (static) logic.

In previous work [14], I begin by proving soundness of exactly this modal logic. I show that this $\mathbf{T}$ modality is sound with respect to some common modal axioms, such as (T) $\mathbf{T}\varphi \rightarrow \varphi$ and (4) $\mathbf{T}\varphi \rightarrow \mathbf{TT}\varphi$. Additionally, I show that $\mathbf{T}$ is a non-normal modality, in the sense that it does *not* satisfy the normal modal axiom (K) $\mathbf{T}(\varphi \wedge \psi) \leftrightarrow (\mathbf{T}\varphi \wedge \mathbf{T}\psi)$. Instead, the modal version of the (Loop) rule serves as a weakening of monotonicity:

$$\text{(Loop)} \quad (\mathbf{T}\varphi_0 \rightarrow \varphi_1 \wedge \ldots \wedge \mathbf{T}\varphi_n \rightarrow \varphi_0) \rightarrow (\mathbf{T}\varphi_0 \rightarrow \varphi_n)$$

In this paper I also used soundness to implement neural network extraction using Tensorflow (albeit for the simplified nets we consider here).

More recently, I have tackled completeness of this logic. I have made significant headway, and I am now one small lemma away from proving completeness of this base modal logic. This result should be unsurprising, since it is again a technical re-write of the previous work on conditionals.

Step 3. Prove soundness and completeness for the logic with $[\varphi]_{\mathbf{hebb}}$
In my previous work [14], I add the Hebbian learning operator $[\varphi]_{\text{hebb}}$ to this base logic. Specifically, this operator follows Hebb's policy: When two adjacent neurons are simultaneously and persistently active, the connection between them strengthens [12]. We implement the naïve (unstable, no weight decay) variant of this policy, i.e. we strengthen connections $W_{i,j}$ according to:

$$\Delta W_{i,j} = \eta x_i x_j$$

where η is the learning rate and x_i, x_j are the binary outputs of adjacent neurons i and j, respectively.

The main contribution of [14] is a proof of soundness for the logic with $[\varphi]_{\text{hebb}}$. I show that $[\varphi]_{\text{hebb}}$ satisfies certain reduction axioms (e.g. $[\varphi]_{\text{hebb}}$ distributes over $\neg$ and $\wedge$). But most interesting is how $[\varphi]_{\text{hebb}}$ interacts with $\mathbf{T}$. We obtain axioms such as

$$[\varphi]_{\text{hebb}}\, \mathbf{T}\psi \to \mathbf{T}[\varphi]_{\text{hebb}}\, \psi$$

which says that if after learning φ, our agent thinks normally ψ, then she would have expected ψ to be true after learning φ in the first place.

Although unpublished, I have recently developed a concrete plan for proving completeness for $[\varphi]_{\text{hebb}}$. First, say we introduce a new modality $\mathbf{K}\varphi$ (read "the agent knows φ") and interpret it as the component of nodes graph-reachable from the neurons for φ. This is a standard move in modal logic (most DELs involve $\mathbf{K}$ in some form). Now consider $[\varphi]^*_{\text{hebb}}$, the transitive closure of the $[\varphi]_{\text{hebb}}$ operator. We have the following axiom that *reduces* $[\varphi]^*_{\text{hebb}}$ to the static logic of $\mathbf{K}$ and $\mathbf{T}$:

$$\begin{aligned} \textbf{(Reduction)} \quad [\varphi]^*_{\text{hebb}}\, \mathbf{T}\psi \leftrightarrow & [(\mathbf{T}\varphi \vee \mathbf{T}\psi \leftrightarrow \top) \wedge \mathbf{T}[\varphi]^*_{\text{hebb}}\, \psi] \\ & \vee [(\mathbf{T}\varphi \vee \mathbf{T}\psi \leftrightarrow \bot) \wedge \mathbf{T}[\varphi]^*_{\text{hebb}}\, \psi \wedge (\mathbf{T}\varphi \vee \mathbf{K}\psi)] \end{aligned}$$

This, along with the following axioms relating $[\varphi]^*_{\text{hebb}}$ to $[\varphi]_{\text{hebb}}$ (see [20])

$$\begin{aligned} &\textbf{(Mix)} && [\varphi]^*_{\text{hebb}}\, \psi \to (\psi \wedge [\varphi]_{\text{hebb}}\, [\varphi]^*_{\text{hebb}}\, \psi) \\ &\textbf{(Induction)} && (\psi \wedge [\varphi]^*_{\text{hebb}}\, (\psi \to [\varphi]_{\text{hebb}})) \to [\varphi]^*_{\text{hebb}}\, \psi \end{aligned}$$

may well be enough to obtain a complete axiomatization of $[\varphi]_{\text{hebb}}$. (If not, then at least we have a strong start.)

As expected, the actual work involved in this completeness proof gives rise to a model-building procedure. I intend to update the Tensorflow implementation to include this.

Step 4. Extend these results to nets with fuzzy activation values
From here, I intend to generalize these results to apply to neural networks used in practice. The first step in this direction is to consider neural networks with fuzzy (real) activation values. But in [10], Giordano et. al. show that this step is just a matter of lifting our rules and axioms from binary logic to fuzzy logic. For example, rather than interpreting $\mathbf{T}\varphi$ as the the crisp set of nodes in the propagation of φ, we can instead interpret it as a *fuzzy set* of activations propagated by φ. We expect most of our axioms to remain intact, although proving that the results transfer may require some new ideas.

Step 5. Extend these results to nets that learn via backpropagation
I expect the story for backpropagation to be more complicated, but at this stage I will have the complete logic of Hebbian learning to guide me. The main idea is that we can introduce the operator $[\varphi; \psi]_{\text{bp}}$ ("learning input φ and its output ψ via one step of backpropagation") and study it the same way we did $[\varphi]_{\text{hebb}}$.

At this point, the code will be of tremendous help with proving completeness — it often happens that writing a program for model-building suggests new axioms that the logic must have.

Evaluation

This work is largely theoretical in nature. My expected research output consists of (1) a series of soundness and completeness proofs relating neural network learning to logic, (2) a code implementation of neural network extraction and model-building using $[\varphi]\psi$-sentences, and (3) a precise philosphical intuition of learning as preference upgrade. The success of (2) and (3) is entirely dependent on the success of (1); the proofs give shape to the *correct* correspondence between neural network learning and dynamic modal logic.

In this sense, (1) is the experiment that tests my thesis. Is it possible to prove such a soundness and completeness result using only off-the-shelf methods from modal logic? Although my thesis is phrased in a deliberately open-ended way, it is in fact falsifiable. It is very possible that there is *no* intuitive, human-readable modal language that completely captures the behavior of neural network learning. And although this seems very unlikely for Hebbian learning, we may have to confront this reality when faced with backpropagation. But I take solace in the fact that if my thesis is false, we can construct a counterexample that teaches us a deep insight about the power of backpropagation.

Timeline

Work Already Completed

Fall 2021.

- Designed and conceptualized the logic of Hebbian learning.
- Proved basic algebraic properties involving forward propagation and Hebbian update.

Spring 2022.

- Proved soundness for the logic of Hebbian learning.
- Implemented (Hebbian) model extraction in Tensorflow.

Summer 2022.

- One small lemma away from proving completeness of base modal logic.
- Discovered reduction axioms for $[\varphi]^*_{\text{hebb}}$ and closure axioms for $[\varphi]_{\text{hebb}}$.

Work Left To Do

Fall 2022.

- Finish up the completeness of the logic of Hebbian learning.
- Begin implementing (Hebbian) model construction in Tensorflow.

Spring 2023.

- Generalize these results for neural networks with fuzzy activation values.
- Make any necessary changes to the Tensorflow implementation.

Summer 2023.

- Design and conceptualize the logic of backpropagation.
- Prove basic properties involving single-step backpropagation.

Fall 2023.

- Prove soundness of the logic of backpropagation.
- Implement (backprop) model extraction in Tensorflow.

Spring 2024.

- Prove completeness of the logic of backpropagation.
- Implement (backprop) model construction in Tensorflow.

Summer 2024.

- This time is reserved for writing up my dissertation, but can also serve as a buffer if I am stuck on the completeness proof.

Fall 2024.

- Finish writing dissertation, and defend!

References

[1] Sebastian Bader and Pascal Hitzler. Dimensions of neural-symbolic integration-a structured survey. *ArXiv preprint cs/0511042*, 2005.

[2] Christian Balkenius and Peter Gärdenfors. Nonmonotonic Inferences in Neural Networks. In *KR*, pages 32–39. 1991.

[3] Alexandru Baltag, Nina Gierasimczuk, Aybüke Özgün, Ana Lucia Vargas Sandoval, and Sonja Smets. A dynamic logic for learning theory. *Journal of Logical and Algebraic Methods in Programming*, 109:100485, 2019.

[4] Alexandru Baltag, Dazhu Li, and Mina Young Pedersen. On the right path: a modal logic for supervised learning. In *International Workshop on Logic, Rationality and Interaction*, pages 1–14. Springer, 2019.

[5] Reinhard Blutner. Nonmonotonic inferences and neural networks. In *Information, Interaction and Agency*, pages 203–234. Springer, 2004.

[6] Léon Bottou. From machine learning to machine reasoning. *Machine learning*, 94(2):133–149, 2014.

[7] Artur S d'Avila Garcez, Krysia Broda, and Dov M Gabbay. Symbolic knowledge extraction from trained neural networks: a sound approach. *Artificial Intelligence*, 125(1-2):155–207, 2001.

[8] Artur S d'Avila Garcez, Luis C Lamb, and Dov M Gabbay. *Neural-symbolic cognitive reasoning*. Springer Science Business Media , 2008.

[9] Jelle Gerbrandy. Dynamic epistemic logic. In *Logic, language and computation, vol. 2*, pages 67–84. 1999.

[10] Laura Giordano, Valentina Gliozzi, and Daniele Theseider Dupré. From common sense reasoning to neural network models through multiple preferences: An overview. *CoRR*, abs/2107.04870, 2021.

[11] Sven Ove Hansson. Logic of belief revision. 2006.

[12] Donald Hebb. *The organization of behavior: A neuropsychological theory*. Wiley, New York, 1949.

[13] The Third AI Summer, AAAI Robert S. Engelmore Memorial Award Lecture. AAAI, 2020.

[14] Caleb Kisby, Saúl Blanco, and Lawrence Moss. The logic of hebbian learning. In *The International FLAIRS Conference Proceedings*, volume 35. 2022.

[15] Luis C Lamb, Artur Garcez, Marco Gori, Marcelo Prates, Pedro Avelar, and Moshe Vardi. Graph neural networks meet neural-symbolic computing: a survey and perspective. *ArXiv preprint arXiv:2003.00330*, 2020.

[16] Hannes Leitgeb. Nonmonotonic reasoning by inhibition nets. *Artificial Intelligence*, 128(1-2):161–201, 2001.

[17] Hannes Leitgeb. Nonmonotonic reasoning by inhibition nets II. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 11(supp02):105–135, 2003.

[18] MartínAbadi et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. 2015. Software available from tensorflow.org.

[19] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.

[20] Lawrence S Moss. Finite models constructed from canonical formulas. *Journal of Philosophical Logic*, 36(6):605–640, 2007.

[21] Jan Plaza. Logics of public communications. *Synthese*, 158(2):165–179, 2007.

[22] Md Kamruzzaman Sarker, Lu Zhou, Aaron Eberhart, and Pascal Hitzler. Neuro-symbolic artificial intelligence: current trends. *ArXiv preprint arXiv:2105.05330*, 2021.

[23] Johan Van Benthem. Dynamic logic for belief revision. *Journal of applied non-classical logics*, 17(2):129–155, 2007.

[24] Johan Van Benthem and Fenrong Liu. Dynamic logic of preference upgrade. *Journal of Applied Non-Classical Logics*, 17(2):157–182, 2007.

[25] Dongran Yu, Bo Yang, Dayou Liu, and Hui Wang. A survey on neural-symbolic systems. *ArXiv preprint arXiv:2111.08164*, 2021.