A Core Method for the Weak Completion Semantics with Skeptical Abduction (Extended Abstract)

Emmanuelle-Anna Dietz Saldanha[1], **Steffen Hölldobler**[1,2*], **Carroline Dewi Puspa Kencana Ramli**[1] and **Luis Palacios Medinacelli**[3,4]
[1]International Center for Computational Logic, Technische Universität Dresden, Germany
[2]North-Caucasus Federal University, Stavropol, Russian Federation
[3]LRASC, Thales Research & Technology, Palaiseau, France
[4]LRI, CNRS, Université Paris-Saclay, France
{dietz, sh}@iccl.tu-dresden.de, carroline.ramli@gmail.com, palacios.medinacelli@gmail.com

Abstract

The Weak Completion Semantics is a novel cognitive theory which has been successfully applied – among others – to the suppression task, the selection task and syllogistic reasoning. It is based on logic programming with skeptical abduction. Each weakly completed program admits a least model under the three-valued Lukasiewicz logic which can be computed as the least fixed point of an appropriate semantic operator. The operator can be represented by a three-layer feed-forward network using the Core method. Its least fixed point is the unique stable state of a recursive network which is obtained from the three-layer feed-forward core by mapping the activation of the output layer back to the input layer. The recursive network is embedded into a novel network to compute skeptical abduction. This extended abstract outlines a fully connectionist realization of the Weak Completion Semantics.

1 Introduction

In his seminal paper on the situation calculus, McCarthy 1963 formulated requirements for systems to reason about actions and causality. Besides being able to specify properties as formulas and to draw conclusions as logical consequences he suggested that *the formal descriptions of states should correspond as closely as possible to what people may reasonably be presumed to know about them when deciding what to do*. In order to meet this latter requirement we need to study humans and their behaviour, which is usually done within Cognitive Science.

In this extended abstract,[1] we are concerned with human reasoning tasks like, e.g., Byrne's [1989] suppression task, Wason's [1968] selection task, or syllogistic reasoning [Khemlani and Johnson-Laird, 2012]. Firstly, we are interested in finding a declarative, computational logic approach adequately modeling these tasks. Secondly, we would like to embed the computational logic approach in a plausible connectionist network.

The Weak Completion Semantics [Hölldobler, 2015] is a new cognitive theory which has been successfully applied to various human reasoning tasks. It is rooted in the work by Stenning and van Lambalgen [2008] but corrects a technical bug by switching from three-valued Kripke-Kleene [Fitting, 1985] to three-valued Łukasiewicz [1920] logic [Hölldobler and Kencana Ramli, 2009].

To illustrate the Weak Completion Semantics consider an example from the suppression task. Suppose we learn that *if she has an essay to write then she will study late in the library*. Under the Weak Completion Semantics the given conditional is represented by the logic program

$$\mathcal{P}_1 = \{\ell \leftarrow e \wedge \neg ab_1, Ab_1 \leftarrow \bot\},$$

where ℓ and e denote that *she will study late in the library* and *she has an essay to write*, respectively, and ab_1 is an abnormality predicate. The weak completion of the program is obtained by adding the only-if halves of the defined relations, i.e., the relations which appear as conclusions in an implication. In our running example, the relations ℓ and ab_1 are defined and, thus, the weak completion of $\mathcal{P}_1$ is the set

$$\{\ell \leftrightarrow e \wedge \neg ab_1,\ ab_1 \leftrightarrow \bot\}$$

of equivalences. One should observe that ab_1 is now mapped to *false*. Moreover, the undefined relation e is not mapped to *false* as it would be under the completion of the program [Clark, 1978].

If we observe that *she will study late in the library*, then according to Byrne [1989], 71% of the participants concluded that *she has an essay to write*. This can be computed in the Weak Completion Semantics by explaining the observation via $\{e \leftarrow \top\}$ given $\mathcal{P}_1$.

Assume additionally, that *if she has a textbook to read then she will study late in the library*. Under the Weak Completion Semantics the two given conditionals are represented by the logic program

$$\mathcal{P}_2 = \mathcal{P}_1 \cup \{\ell \leftarrow t \wedge \neg ab_2,\ ab_2 \leftarrow \bot\},$$

*Contact Author

[1]This is an extended abstract of an article entitled *A Core Method for the Weak Completion Semantics with Skeptical Abduction* published in the Journal of Artificial Intelligence Research [Dietz Saldanha *et al.*, 2018].

where t denotes that *she has a textbook to read*, and ab_2 is another abnormality predicate, which is also assumed to be false. If we observe again that *she will study late in the library*, then the previously drawn conclusion is suppressed in that only 13% of the participants concluded that *she has an essay to write* [Byrne, 1989]. If we apply abduction under the Weak Completion Semantics then the observation can be explained by the two minimal explanations $\{e \leftarrow \top\}$ and $\{t \leftarrow \top\}$. Hence, reasoning credulously we would conclude that *she has an essay to write*. Apparently, humans don't do this; they appear to reason skeptically.

Programs like $\mathcal{P}_1$ and $\mathcal{P}_2$ as well as their weak completions admit a least model under Łukasiewicz logic which can be computed as the least fixed point of an appropriate semantic operator [Hölldobler and Kencana Ramli, 2009]. In this context, we are using the operator $\Phi_\mathcal{P}$ for a logic program $\mathcal{P}$ which was introduced by Stenning and van Lambalgen [2008].

Semantic operators for logic programs were first studied by Apt and van Emden [1982] in an attempt to capture the semantics of logic programs. In Cognitive Science, semantic operators are interesting as they allow to construct a model for a given program. This construction may be compared to other ways of generating models like, for example, in the theory of mental models [Johnson-Laird and Byrne, 1991]. Suppose we extend the program $\mathcal{P}_1$ with the fact $e \leftarrow \top$ representing the information that *she has an essay to write*. The extended program corresponds to the case of modus ponens in the suppression task [Byrne, 1989]. Starting with the empty interpretation, i.e., the interpretation where all relations are unknown, the semantic operator of the Weak Completion Semantics assigns *true* to e and *false* to ab_1 in its first application because of the positive fact $e \leftarrow \top$ and the negative assumption $ab_1 \leftarrow \bot$, respectively. In its second application, the operator additionally assigns *true* to ℓ because the condition of the rule

$$\ell \leftarrow e \wedge \neg ab_1$$

is *true* as soon as e is mapped to *true* and ab_1 is mapped to *false*. In other words, ℓ being *true* is an immediate consequence of the rule given that its condition is *true*. Further applications of the semantic operator do not alter the findings. A least fixed point has been reached. Reasoning with respect to this least fixed point allows to conclude ℓ, which is what 96% of the subjects in the suppression task do.

The semantic operator introduced by Apt and van Emden [1982] is continuous. Funahashi [1989] has shown that continuous mappings can be approximated arbitrary well by feed-forward networks. Combining both results, Hölldobler and Kalinke [1994] developed the idea to compute semantic operators for propositional logic programs by feed-forward networks. By connecting the output to the input layer, the feed-forward networks are turned into recurrent ones. These recurrent networks compute iterated applications of the semantic operators and, in particular, if they reach a stable state, then this state corresponds to the least fixed point of the semantic operator. In other words, the connectionist networks compute the least models of the given programs. The idea was later extended to first-order programs and called CORE method for connectionist model generation using recurrent networks with feed-forward core [Bader and Hölldobler, 2006].

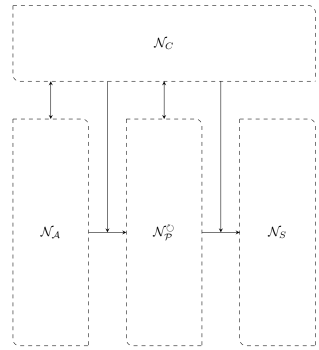

Figure 1: A schematic view of the network.

The first question to be considered in this paper is: *How can the semantic operator associated with the Weak Completion Semantics be represented and computed within a fully connectionist setting?* An extension of the CORE method to three-valued Łukasiewicz [1920] logic is not enough to answer this question. We additionally need to extend the network to determine

- whether it has reached a stable state,
- to consider additional facts and assumptions in order to explain a given observation,
- to check whether a given set of integrity constraints is satisfied and
- to eliminate stable coalitions of units that may have arisen from previous reasoning episodes.

However, even with such a network we cannot solve all reasoning episodes of the suppression task. As discussed before, we have to add skeptical abduction. Unfortunately, all known connectionist solutions which we are aware of handle credulous abduction in classical two-valued logic [Ray and d'Avila Garcez, 2006; d'Avila Garcez *et al.*, 2007]. Hence, the second question to be considered in this paper is: *How can skeptical abduction be integrated into the connectionist realization of the* CORE *method?* In particular, we will use McCulloch-Pitts [1943] networks to sequentially generate all possible explanations. Possible explanations are forwarded to the CORE network. As soon as an explanation is detected, it will be stored. Once all possible explanations are tested, the skeptical conclusions will be computed.

2 A Schematic View of the Network

Given is a logic program $\mathcal{P}$, its set $\mathcal{A}$ of abducibles consists of all facts of the form $A \leftarrow \top$ and assumptions $A \leftarrow \bot$ of atoms A which are undefined in $\mathcal{P}$. Let $\mathcal{O}$ be an observation, i.e., a set of literals, and $\mathcal{IC}$ be a set of integrity constraints. The goal of the network is to compute the skeptical consequences of $\mathcal{P}$ while explaining $\mathcal{O}$ and satisfying $\mathcal{IC}$.

In Figure 1 a schematic view of the connectionist network for $\mathcal{P}$, $\mathcal{A}$, $\mathcal{O}$ and $\mathcal{IC}$ is depicted. It can be divided into four subnetworks:

- $\mathcal{N}_\mathcal{A}$: a network to sequentially generate all possible explanations given the set $\mathcal{A}$ of abducibles.
- $\mathcal{N}_\mathcal{P}^\circlearrowleft$: a recursive network to compute the least model of the semantic operator associated with the weak completion of the program $\mathcal{P}$ and a possible explanation.
- $\mathcal{N}_S$: a network to compute the skeptical consequences.
- $\mathcal{N}_C$: a network to control the process.

The subnetworks are sketched in the following sections. They are specified in full detail in [Dietz Saldanha *et al.*, 2018].

3 Generating Possible Explanations

For a finite (propositional or datalog) program $\mathcal{P}$ the set $\mathcal{A}$ of abducibles is also finite. The task is to generate a sequence of all possibles explanations, i.e., all subsets of $\mathcal{A}$. This can be formally specified as a finite automaton with state output (Moore machine). McCulloch and Pitts [1943] have shown that each finite automaton with state output can be turned into a connectionist network of binary threshold units receiving an input via its input units and producing an output via its output units. $\mathcal{N}_\mathcal{A}$ is such a network. Whenever prompted by $\mathcal{N}_C$ it produces the next possible explanation $\mathcal{E}$ and forwards it to $\mathcal{N}_\mathcal{P}^\circlearrowleft$. As soon as all possible explanations have been generated, $\mathcal{N}_\mathcal{A}$ informs $\mathcal{N}_C$. The network has been improved in the meantime (see Section 7).

4 Computing Least Models

For a finite (propositional or datalog) program $\mathcal{P}$ the semantic operator $\Phi_\mathcal{P}$ introduced by Stenning and van Lambalgen [2008] is continuous. Hence, it can be computed by a feed-forward network based on the ideas originally developed in [Hölldobler and Kalinke, 1994]. The network can also be trained via backpropagation based on ideas originally developed in [d'Avila Garcez *et al.*, 1997].

The core of the network $\mathcal{N}_\mathcal{P}^\circlearrowleft$ is a feed-forward network with an input, a hidden and an output layer. It computes the semantic operator $\Phi_{\mathcal{P}\cup\mathcal{E}}$ for a program $\mathcal{P}$ together with a possible explanation $\mathcal{E}$. $\mathcal{E}$ has been computed by $\mathcal{N}_\mathcal{A}$ before (see Section 3) and is submitted to $\mathcal{N}_\mathcal{P}^\circlearrowleft$ by clamping units occurring in $\mathcal{N}_\mathcal{P}^\circlearrowleft$ which are representing $\mathcal{E}$. The feed-forward core is turned into a recurrent network by connecting the units of the output layer to the units of the input layer such that $\Phi_{\mathcal{P}\cup\mathcal{E}}$ is applied recursively until a stable state is reached. Such a stable state always exists. It corresponds to the least fixed point of $\Phi_{\mathcal{P}\cup\mathcal{E}}$ and the least model of the weak completion of $\mathcal{P}\cup\mathcal{E}$.

Once a stable state has been reached, the control network $\mathcal{N}_C$ checks whether the stable state explains the observation $\mathcal{O}$ and satisfies the integrity constraints $\mathcal{IC}$. If this is the case, then the least model of the weak completion of $\mathcal{P}\cup\mathcal{E}$ is propagated from $\mathcal{N}_\mathcal{P}^\circlearrowleft$ to $\mathcal{N}_S$. At the same time, the explanation $\mathcal{E}$ is withdrawn. As the external clamping of units occurring in $\mathcal{N}_\mathcal{P}^\circlearrowleft$ may lead to stable coalitions of units which persist even if the external activation is withdrawn, stable coalitions occurring in the network $\mathcal{N}_\mathcal{P}^\circlearrowleft$ must be de-activated before the next possible explanation can be considered. This is achieved by inhibiting the input units of $\mathcal{N}_\mathcal{P}^\circlearrowleft$ for three time steps. Thereafter, all units of $\mathcal{N}_\mathcal{P}^\circlearrowleft$ are passive.

5 Computing Skeptical Conclusions

The network $\mathcal{N}_S$ is a two-layer feed-forward network which combines the least models of the weak completion of $\mathcal{P}\cup\mathcal{E}$ for all explanations $\mathcal{E}$ for the observation $\mathcal{O}$ which are satisfying the integrity constraints $\mathcal{IC}$. For each atom A, the network has three input units representing its possible values under Łukasiewicz logic: *true*, *false* or *unknown*. The units in the input layer are self-excitatory. Hence, least models generated by $\mathcal{N}_\mathcal{P}^\circlearrowleft$ are stored and persist. The activation of the unit representing that atom A is mapped to *true* is propagated to the corresponding unit in the output layer if all least models generated by $\mathcal{N}_\mathcal{P}^\circlearrowleft$ for the various explanations of $\mathcal{O}$ map A to *true*. Likewise, the activation of the unit representing that atom A is mapped to *false* is propagated to the corresponding unit in the output layer if all least models generated by $\mathcal{N}_\mathcal{P}^\circlearrowleft$ map A to *false*. Otherwise, A is mapped to *unknown*.

6 Control

The control network $\mathcal{N}_C$ controls the process of computing skeptical consequences. In particular:

1. It prompts $\mathcal{N}_\mathcal{A}$ to generate the next possible explanation until $\mathcal{N}_\mathcal{A}$ signals that all possible explanations have been generated. Thereafter, the computation is terminated.
2. As soon as the next possible explanation is generated, it allows $\mathcal{N}_\mathcal{A}$ to propagate the possible explanation $\mathcal{E}$ to $\mathcal{N}_\mathcal{P}^\circlearrowleft$ by clamping corresponding units occurring in $\mathcal{N}_\mathcal{P}^\circlearrowleft$.
3. It checks whether $\mathcal{N}_\mathcal{P}^\circlearrowleft$ has reached a stable state upon which it further checks whether the observation $\mathcal{O}$ is explained and the integrity constraints $\mathcal{IC}$ is satisfied.
4. If $\mathcal{O}$ is explained and $\mathcal{IC}$ is satisfied then it allows $\mathcal{N}_\mathcal{P}^\circlearrowleft$ to propagate the least model to $\mathcal{N}_S$, it withdraws the current possible explanation from $\mathcal{N}_\mathcal{P}^\circlearrowleft$ and it forces $\mathcal{N}_\mathcal{P}^\circlearrowleft$ to de-activate stable coalitions.

7 Future Work

Recall from the introduction, that in order to meet McCarthy's 1963 requirements for systems to reason about actions and causalities, we need to study humans and their behavior. It has been previously shown that skeptical abduction is required in order to adequately model a wide range of human reasoning tasks under the Weak Completion Semantics. Motivated by this observation, the main goal of this paper was to specify a plausible connectionist realization of the Weak Completion Semantics, that provides the representation of logic programs, the computation of least models, the interpretation under three-valued Łukasiewicz logic and the derivation of consequences under skeptical abduction.

Lourêdo Rocha [2017] has added minimality to the approach presented in this paper in that only minimal explanations are considered. Moreover, she has replaced the McCulloch-Pitts networks used to generate all possible explanations in a fixed and pre-defined sequence by Elman [1989; 1990] networks and showed that these networks can be trained to generate all possible explanations in an arbitrary order.

This is a prerequisite for the next step: We do not believe that humans test all possible explanations in a systematic way. From a complexity point of view, skeptical conclusions is exponential in the number of abducibles. If reasoning tasks and, in particular, the sets of abducibles considered in abductive reasoning tasks become larger, then it seems unlikely that humans consider all possible explanations. It appears to us that in particular reasoning episodes some possible explanations are systematically tested whereas others are not considered at all. We believe that there is a kind of attention formalism which identifies those possible explanations, which are really tested. This will lead to a kind of bounded skeptical abduction. But it remains to set up experiments that test this hypothesis and, if our hypothesis is supported, then we need to identify the mechanism which defines the bound and build it into our networks.

References

[Apt and van Emden, 1982] Krzysztof R. Apt and Maarten H. van Emden. Contributions to the theory of logic programming. *Journal of the ACM*, 29:841–862, 1982.

[Bader and Hölldobler, 2006] Sebastian Bader and Steffen Hölldobler. The Core method: Connectionist model generation. In *Proceedings of the 16th International Conference on Artificial Neural Networks*, volume 4132 of *Lecture Notes in Computer Science*, pages 1–13. Springer-Verlag, 2006.

[Byrne, 1989] Ruth M. J. Byrne. Suppressing valid inferences with conditionals. *Cognition*, 31:61–83, 1989.

[Clark, 1978] Keith L. Clark. Negation as failure. In Herv/'e Gallaire and Jack Minker, editors, *Logic and Databases*, pages 293–322. Plenum, New York, 1978.

[d'Avila Garcez *et al.*, 1997] Artur S. d'Avila Garcez, Gerson Zaverucha, and Luís A. V. de Carvalho. Logic programming and inductive learning in artificial neural networks. In Christoph Herrmann, Frank Reine, and Antje Strohmaier, editors, *Knowledge Representation in Neural Networks*, pages 33–46, Berlin, 1997. Logos Verlag.

[d'Avila Garcez *et al.*, 2007] Artur S. d'Avila Garcez, Dov M. Gabbay, Oliver Ray, and John Woods. Abductive reasoning in neural-symbolic learning systems. *Topoi: An International Review of Philosophy*, 26:37–49, 2007.

[Dietz Saldanha *et al.*, 2018] Emmanuelle-Anna Dietz Saldanha, Steffen Hölldobler, Carroline D. P. Kencana Ramli, and Luis Palacios Medinacelli. A Core method for the weak completion semantics with skeptical abduction. *Journal of Artificial Intelligence Research*, 63:51–86, 2018.

[Elman, 1989] Jeffrey L. Elman. Structured representations and connectionist models. In *Proceedings of the Annual Conference of the Cognitive Science Society*, pages 17–25, 1989.

[Elman, 1990] Jeffrey L. Elman. Finding structure in time. *Cognitive Science*, 14:179–211, 1990.

[Fitting, 1985] Melvin Fitting. A Kripke–Kleene semantics for logic programs. *Journal of Logic Programming*, 2(4):295–312, 1985.

[Funahashi, 1989] Ken-Ichi Funahashi. On the approximate realization of continuous mappings by neural networks. *Neural Networks*, 2:183–192, 1989.

[Hölldobler and Kalinke, 1994] Steffen Hölldobler and Yvonne Kalinke. Towards a new massively parallel computational model for logic programming. In *Proceedings of the ECAI94 Workshop on Combining Symbolic and Connectionist Processing*, pages 68–77. ECCAI, 1994.

[Hölldobler and Kencana Ramli, 2009] Steffen Hölldobler and Carroline D. P. Kencana Ramli. Logic programs under three-valued Łukasiewicz's semantics. In *Logic Programming*, volume 5649 of *Lecture Notes in Computer Science*, pages 464–478. Springer-Verlag, 2009.

[Hölldobler, 2015] Steffen Hölldobler. Weak completion semantics and its applications in human reasoning. In Ulrich Furbach and Claudia Schon, editors, *Bridging 2015 – Bridging the Gap between Human and Automated Reasoning*, volume 1412 of *CEUR Workshop Proceedings*, pages 2–16. CEUR-WS.org, 2015. http://ceur-ws.org/Vol-1412/.

[Johnson-Laird and Byrne, 1991] Philip N. Johnson-Laird and Ruth M. J. Byrne. *Deduction*. Lawrence Erlbaum Associates, Hove and London (UK), 1991.

[Khemlani and Johnson-Laird, 2012] Sangeet Khemlani and Philip N. Johnson-Laird. Theories of the syllogism: A meta-analysis. *Psychological Bulletin*, 138(3):427–457, 2012.

[Lourêdo Rocha, 2017] Isabelly Lourêdo Rocha. Bounded sceptical reasoning. Master's thesis, International Center for Computational Logic, TU Dresden, 2017.

[Łukasiewicz, 1920] Jan Łukasiewicz. O logice trójwartościowej. *Ruch Filozoficzny*, 5:169–171, 1920. English translation: On Three-Valued Logic. In: *Jan Łukasiewicz Selected Works*. (Ludwik Borkowski, ed.), North Holland, 87-88, 1990.

[McCarthy, 1963] John McCarthy. Situations and actions and causal laws. Stanford Artificial Intelligence Project: Memo 2, 1963.

[McCulloch and Pitts, 1943] Warren S. McCulloch and Walter Pitts. A logical calculus and the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:115–133, 1943.

[Ray and d'Avila Garcez, 2006] Oliver Ray and Artur d'Avila Garcez. Towards the integration of abduction and induction in artificial neural networks. In *Proceedings of the ECAI06 Workshop on Neural-Symbolic Learning and Reasoning*, pages 41–46, 2006.

[Stenning and van Lambalgen, 2008] Keith Stenning and Michiel van Lambalgen. *Human Reasoning and Cognitive Science*. MIT Press, 2008.

[Wason, 1968] Peter C. Wason. Reasoning about a rule. *The Quarterly Journal of Experimental Psychology*, 20:273–281, 1968.