Activation-Based Conditional Inference (Extended Abstract)*

Marco Wilhelm[1], Diana Howey[1], Gabriele Kern-Isberner[1], Kai Sauerwald[2], and Christoph Beierle[2]

[1] Dept. of Computer Science, TU Dortmund University, Dortmund, Germany
[2] Faculty of Mathematics and Computer Science, FernUniversität in Hagen, Hagen, Germany

1 Introduction

Activation-based conditional inference (ActInf) [12] combines conditional reasoning and *ACT-R (Adaptive Control of Thought-Rational)* [3, 2], an architecture from cognitive science used to formalize human reasoning, and therewith constitutes a powerful and cognitively well-founded inference framework. Basically, there are two cognitive processes in human reasoning which affect the selection of beliefs (called *chunks* in ACT-R): The long-term process of forgetting and remembering and the short-term process of activating specific beliefs depending on the context. ACT-R formalizes these processes in form of an *activation function* which models the retrieval of beliefs. Based on this selection of beliefs, the inference in typical ACT-R implementations is realized by a *production system* [5, 7]. ActInf reinterprets the basic elements of ACT-R in a more abstract form, detaching them from its inference engine using production rules. In this way, ActInf makes ACT-R more broadly usable for knowledge representation. By combining the abstract ACT-R elements with state-of-the-art conditional reasoning operators, activation-based conditional inference presents as a novel approach to reasoning that explicitly takes cognitive aspects into account.

2 Retrieval of Beliefs in ACT-R

In *ACT-R* [3, 2] a distinction is made between the *declarative memory* in which categorical knowledge about individuals or objects is stored in form of *chunks (knowing that)* and the *procedural memory* which typically consists of *production rules* and describes how the chunks are processed *(knowing how)* [10]. Reasoning in ACT-R starts with an *initial priming* which causes an activation of chunks. The chunk with the highest activation is selected and processed by production rules in order to compute a solution of the reasoning task. If this fails, the activation passes into an iterative process: The system obtains additional chunks from the declarative memory and tries to compute a solution again until either the problem is solved or no further chunks are active.

* This introduction to activation-based conditional inference is based on [12].

The retrieval of chunks ($\mathfrak{c}_i$) is formalized by an *activation function* $\mathcal{A}(\mathfrak{c}_i)$ which is calculated for each specific request anew and which is based on a *usage history* of the chunks, associations between *cognitive units* and the *priming* [1]. It is given as the sum of the so-called *base-level activation* $\mathcal{B}(\mathfrak{c}_i)$ and the *spreading activation* $\mathcal{S}(\mathfrak{c}_i)$, which again is a sum of *degrees of association* between chunks $\mathcal{S}(\mathfrak{c}_i, \mathfrak{c}_j)$ weighted by a *weighting factor* $\mathcal{W}(\mathfrak{c}_j)$:

$$\mathcal{A}(\mathfrak{c}_i) = \underbrace{\mathcal{B}(\mathfrak{c}_i)}_{\text{base-level activation}} + \underbrace{\sum\nolimits_j \mathcal{W}(\mathfrak{c}_j) \cdot \mathcal{S}(\mathfrak{c}_i, \mathfrak{c}_j)}_{\text{spreading activation } \mathcal{S}(\mathfrak{c}_i)} . \tag{1}$$

The *base-level activation* of a chunk $\mathcal{B}(\mathfrak{c}_i)$ reflects the *entrenchment* of $\mathfrak{c}_i$ in the reasoner's memory and depends on the recency and frequency of its use. Therewith, it models the long-term memory of the reasoner. The *spreading activation* describes the context-dependency of the activation of chunks, hence, the short-term process. The core idea behind the *spreading activation theory* [1] is that an initial priming caused by sensory stimuli triggers certain *cognitive units*, the basic building blocks of chunks, which again trigger related cognitive units and so on until the disposition for activation is too low to trigger further cognitive units. Usually, the cognitive units are arranged as vertices in an undirected graph, the so-called *spreading activation network*, and the triggering of cognitive units is propagated through this network. The spreading activation $\mathcal{S}(\mathfrak{c}_i)$ can then be derived from the *triggering values* of the cognitive units of which $\mathfrak{c}_i$ makes use. The interrelation of cognitive units and of chunks is specified in more detail in the *degree of association* and the *weighting factor*. The *degree of association* $\mathcal{S}(\mathfrak{c}_i, \mathfrak{c}_j)$ reflects how strongly related $\mathfrak{c}_i$ and $\mathfrak{c}_j$ are. Chunks which deal with the same issue (i.e., share cognitive units) have a high degree of association while chunks which refer to different topics are only loosely or not related and therefore have a low degree of association. While the degree of association is independent of the priming, the weighting factor reflects the context-dependency of the spreading activation.

3 Activation-Based Conditional Inference

Activation-based conditional inference (ActInf) [12] proposes a cognitively inspired model of inductive conditional reasoning by interpreting the concepts of ACT-R in terms of logic, conditionals, and inference. In ActInf, the declarative memory is formalized as a conditional belief base Δ so that chunks are replaced by conditionals $\delta \in \Delta$. Conditionals are expressions of the form $(B|A)$, here defined over a propositional language $\mathcal{L}(\Sigma)$ with $A, B \in \mathcal{L}(\Sigma)$, stating that "$B$ usually follows from A." Based on an adapted version of the activation function (1) a subset of Δ is computed from which inferences are drawn with respect to any *inductive inference operator* [4] which replaces the production system of ACT-R. If $\mathfrak{q}$ is a query conditional, which serves as the initial priming in ActInf, the activation function for conditionals is given by

$$\mathcal{A}_{\mathfrak{q}}^{\Delta}(\delta) = \underbrace{\mathcal{B}^{\Delta}(\delta)}_{\text{base-level activation}} + \underbrace{\sum_{\delta' \in \Delta} \mathcal{W}_{\mathfrak{q}}^{\Delta}(\delta') \cdot \mathcal{S}(\delta, \delta')}_{\text{spreading activation } \mathcal{S}_{\mathfrak{q}}^{\Delta}(\delta)}, \qquad \delta \in \Delta.$$

In order to answer $\mathfrak{q}$, for every conditional in Δ its activation value $\mathcal{A}_{\mathfrak{q}}^{\Delta}(\delta)$ is calculated. If this value is above a certain threshold, then the conditional is selected for drawing the inference. The iterative retrieval of chunks in ACT-R can be mimicked in this framework by iteratively lowering the threshold. Note that one has to be mindful of the non-monotonicity of conditional reasoning and the three-valued logic of conditionals in this context. A way of guaranteeing that during the iteration process formerly drawn inferences are not rejected in the further course is by choosing a semi-monotonous inference operator (like *System P inference* [6], cf. also the connection to *focused inference* [13]) and considering a query to fail if neither the query conditional $\mathfrak{q} = (B|A)$ nor its counterpart $(\neg B|A)$ can be inferred.

In the remainder, we recall the ideas from [12] on how the components of $\mathcal{A}_{\mathfrak{q}}^{\Delta}(\delta)$ can be configured in order to obtain an executable activation function which can be seen as a blueprint for the development of further ActInf systems.

Base-Level Activation. The base-level activation $\mathcal{B}^{\Delta}(\delta)$ reflects the entrenchment of δ in the reasoner's memory. Since epistemic entrenchment and *ranking semantics* [11] are dual ratings, the *normality* of a conditional is a good estimator. The normality of a conditional δ can be expressed by its *Z-rank* $Z^{\Delta}(\delta)$ specified by *System Z* [9] so that we define

$$\mathcal{B}^{\Delta}(\delta) = \frac{1}{1 + Z^{\Delta}(\delta)}, \qquad \delta \in \Delta.$$

While the most normal conditionals have a base-level activation of $\mathcal{B}^{\Delta}(\delta) = 1$, this value decreases with increasing specificity of δ.

Weighting Factor. The weighting factor $\mathcal{W}_{\mathfrak{q}}^{\Delta}(\delta)$ indicates how much the initial priming $\mathfrak{q}$ triggers the conditional δ. We formalize the influence of the priming according to the spreading activation theory by a labeling of a *spreading activation network* $\mathcal{N}(\Delta) = (\mathcal{V}, \mathcal{E})$, i.e., an undirected graph between cognitive units. In our context, the cognitive units are the atoms $\mathfrak{a} \in \Sigma$ and the outcome of the labeling of $\mathcal{N}(\Delta)$ are *triggering values* $\tau_{\mathfrak{q}}^{\Delta}(\mathfrak{a}) \in [0, 1]$ which indicate how much $\mathfrak{a}$ is triggered by $\mathfrak{q}$. Edges in $\mathcal{N}(\Delta)$ represent associations between the atoms in Σ along which the triggering of the atoms spreads. Two atoms are considered to be associated if they occur commonly in some conditionals in Δ, so that

$$\mathcal{V} = \Sigma \quad \text{and} \quad \mathcal{E} = \{\{\mathfrak{a}, \mathfrak{b}\} \mid \exists \delta \in \Delta : \{\mathfrak{a}, \mathfrak{b}\} \subseteq \Sigma(\delta)\}.$$

The actual spreading of activation is modeled by iteratively labeling the vertices (atoms) in $\mathcal{N}(\Delta)$ with their triggering value $\tau_{\mathfrak{q}}^{\Delta}(\mathfrak{a})$. The *labeling algorithm* is

Input: Spreading activation network $\mathcal{N}(\Delta) = (\mathcal{V}, \mathcal{E})$, query $\mathfrak{q} = (B|A)$
Output: Labeling of $\mathcal{N}(\Delta)$, i.e., triggering values $\tau_{\mathfrak{q}}^{\Delta}(\mathfrak{a}) = \mathit{label}(\mathfrak{a})$ for $\mathfrak{a} \in \Sigma$

for $\mathfrak{a} \in \mathcal{V}$ with $\mathfrak{a} \in \Sigma(\mathfrak{q})$: $\mathit{label}(\mathfrak{a}) = 1$
initialize
    $\mathcal{L} = \{\mathfrak{a} \in \mathcal{V} \mid \mathfrak{a} \text{ is labeled}\}$,
    $\mathcal{V}' = \{\mathfrak{a} \in \mathcal{V} \mid \exists\{\mathfrak{a}, \mathfrak{b}\} \in \mathcal{E}\colon \mathfrak{a} \in \mathcal{V} \setminus \mathcal{L} \wedge \mathfrak{b} \in \mathcal{L}\}$
while $\mathcal{V}' \neq \emptyset$:
    for $\mathfrak{a} \in \mathcal{V}'$: $\mathit{label}(\mathfrak{a}) = \dfrac{\sum_{\mathfrak{b} \in \mathcal{L}\colon \{\mathfrak{a},\mathfrak{b}\} \in \mathcal{E}} \mathit{label}(\mathfrak{b})}{1 + \sum_{\mathfrak{b} \in \mathcal{L}} \mathit{label}(\mathfrak{b})}$
    update $\mathcal{L}$, $\mathcal{V}'$
for $\mathfrak{a} \in \mathcal{V} \setminus \mathcal{L}$: $\mathit{label}(\mathfrak{a}) = 0$
return $\mathit{label}(\mathfrak{a})$ for $\mathfrak{a} \in \mathcal{V}$

Fig. 1. Labeling algorithm.

shown in Figure 1 and an example of a spreading activation network together with its labels is shown in Figure 2.

Eventually, we follow the idea that a conditional δ cannot be triggered more than its atoms and define the weighting factor by

$$\mathcal{W}_{\mathfrak{q}}^{\Delta}(\delta) = \min\{\tau_{\mathfrak{q}}^{\Delta}(\mathfrak{a}) \mid \mathfrak{a} \in \Sigma(\delta)\}, \qquad \delta \in \Delta.$$

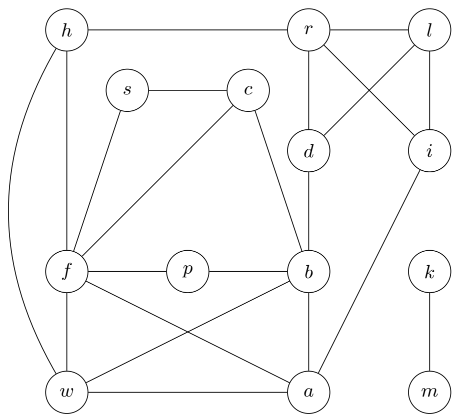
Degree of Association. The degree of association $\mathcal{S}(\delta, \delta')$ is a measure of connectedness between the conditionals in Δ and is defined by

$$S(\delta, \delta') = \frac{|\Sigma(\delta) \cap \Sigma(\delta')|}{|\Sigma(\delta) \cup \Sigma(\delta')|}, \qquad \delta, \delta' \in \Delta.$$

It is the number of shared atoms relative to all atoms in δ or δ'. Note that not only the quantities $\mathcal{S}(\delta, \delta')$ for $\delta' \in \Delta$ themselves determine the spreading activation of a conditional δ but also how *many* conditionals δ is associated with. The more a conditional is cross-linked within Δ, the more likely it is that this conditional has a high spreading activation. With increasing syntactically linkage to the query $\mathfrak{q}$ the spreading activation of a conditional increases, while conditionals which are not linked to the query at all have a spreading activation of zero. In this sense, ActInf refines *syntax splitting* [8, 4] by a gradual notion. The base-level activation "softens" the syntactically motivated rejection of conditionals if the entrenchment of these conditionals is high enough.

4 Forgetting and Remembering in ActInf

Eventually, we present how cognitive concepts like forgetting and remembering can be realized in ActInf. In ACT-R the base-level activation of a chunk is not constant but decreases over time and increases when the chunk is retrieved which we have not considered so far in ActInf. In order to capture this dynamic view on



$\mathfrak{a}$	$\tau^{\Delta}_{\mathfrak{q}_1}(\mathfrak{a})$		$\tau^{\Delta}_{\mathfrak{q}_2}(\mathfrak{a})$	
a	1	(0)	$1/4$	(1)
b	$2/3$	(1)	$1/4$	(1)
c	$4/15$	(2)	1	(0)
d	$2/15$	(2)	$1/21$	(2)
f	$2/3$	(1)	1	(0)
h	$1/5$	(2)	$1/4$	(1)
i	$1/3$	(1)	$1/21$	(2)
k	0	(∞)	0	(∞)
l	$1/15$	(2)	$4/151$	(3)
m	0	(∞)	0	(∞)
p	1	(0)	$1/4$	(1)
r	$1/15$	(2)	$1/21$	(2)
s	$2/15$	(2)	1	(0)
w	$1/3$	(1)	$1/4$	(1)

Fig. 2. Example of a spreading activation network $\mathcal{N}(\Delta)$ for $\Delta = \{(f|aw), (\neg f|a \wedge \neg w), (b \Rightarrow a|\top), (w|b), (d|b), (p \Rightarrow b|\top), (\neg f|p), (c \Rightarrow b|\top), (\neg f|c), (f|c \wedge s), (\neg s|c), (i \Rightarrow a|\top), (r \Rightarrow i|\top), (l \Rightarrow i|\top), (l \vee r|i), (\neg d|r), (d|l), (h \Rightarrow r|\top), (f \wedge \neg w|h), (k \Rightarrow m|\top)\}$ (cf. [12]) and its labels wrt. the queries $\mathfrak{q}_1 = (p \Rightarrow a|\top)$ and $\mathfrak{q}_2 = (f|c \wedge \neg s)$. The numbers in the parentheses next to the labels are the iteration steps in which the atoms are labeled. 0 stands for the priming and ∞ for unreachable atoms.

the base-level activation, we decrease the base-level activation of a conditional when the conditional is not selected for answering a query ($s(\delta) = 0$) and increase it otherwise ($s(\delta) = 1$). Let $\theta \geq 0$. Then, the so-called *forgetting factor* of δ is

$$\phi_{\theta,s}(\delta) = \begin{cases} 1+\theta & \text{iff} \quad s(\delta) = 1 \\ 1-\theta & \text{otherwise} \end{cases}, \qquad \delta \in \Delta.$$

After performing an inference task, we update the base-level activation with this forgetting factor and obtain for the updated base-level activation

$$\mathcal{B}^{\Delta}_{\theta,s}(\delta) = \mathcal{B}^{\Delta}(\delta) \cdot \phi_{\theta,s}(\delta).$$

The higher the parameter θ, the greater the impact of the forgetting factor $\phi_{\theta,s}(\delta)$ on the base-level activation is. When applying this kind of update to a series of inference requests, the usage history of the conditionals is implemented into the base-level activation. Now, we can understand *forgetting* as the process of lowering the base-level activation of a conditional δ so far over time (because it has not been selected for inference requests in the meantime) that δ is not considered for answering a query $\mathfrak{q}$ anymore although it would have been considered if $\mathfrak{q}$ was asked initially. *Remembering* in ActInf can be understood as follows: It can happen that a conditional is not selected for answering a query $\mathfrak{q}$ but over time this conditional is considered in other reasoning tasks such that its base-level activation increases to such an extend that when it comes to the query $\mathfrak{q}$ again this conditional is selected as of now.

Acknowledgments. This work was supported by DFG Grant KE 1413/12-1 awarded to Gabriele Kern-Isberner and DFG Grant BE 1700/10-1 awarded to Christoph Beierle as part of the priority program "Intentional Forgetting in Organizations" (SPP 1921).

References

1. Anderson, J.R.: A spreading activation theory of memory. Journal of Verbal Learning and Verbal Behavior **22**, 261–295 (1983)
2. Anderson, J.R.: How can the human mind occur in the physical universe? Oxford University Press (2007)
3. Anderson, J.R., Lebiere, C.: The atomic components of thought. Psychology Press (1998)
4. Kern-Isberner, G., Beierle, C., Brewka, G.: Syntax splitting = relevance + independence: New postulates for nonmonotonic reasoning from conditional belief bases. In: Calvanese, D., Erdem, E., Thielscher, M. (eds.) Proceedings of the 17th International Conference on Principles of Knowledge Representation and Reasoning, KR 2020. pp. 560–571 (2020)
5. Klahr, D., Langley, P., Neches, R. (eds.): Production System Models of Learning and Development. MIT Press (1987)
6. Kraus, S., Lehmann, D., Magidor, M.: Nonmonotonic reasoning, preferential models and cumulative logics. Artif. Intell. **44**(1-2), 167–207 (1990)
7. Newell, A.: Unified Theories of Cognition. Harvard University Press (1990)
8. Parikh, R.: Beliefs, Belief Revision, and Splitting Languages, pp. 266–278. Center for the Study of Language and Information (1999)
9. Pearl, J.: System Z: A natural ordering of defaults with tractable applications to nonmonotonic reasoning. In: Parikh, R. (ed.) Proceedings of the 3rd Conference on Theoretical Aspects of Reasoning about Knowledge. pp. 121–135. Morgan Kaufmann (1990)
10. Ryle, G.: The Concept of Mind. University of Chicago Press, New edn. (2000)
11. Spohn, W.: The Laws of Belief: Ranking Theory and Its Philosophical Applications. Oxford University Press (2012)
12. Wilhelm, M., Howey, D., Kern-Isberner, G., Sauerwald, K., Beierle, C.: Activation-based conditional inference. Journal of Applied Logics **10**(2) (2023)
13. Wilhelm, M., Kern-Isberner, G.: Focused inference and System P. In: Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021. pp. 6522–6529. AAAI Press (2021)