Topological Aspects of First-Order Neural-Symbolic Integration

Pascal Hitzler

AIFB

Universität Karlsruhe (TH)

Invited talk for the lecture on Logic Programming and Connectionist Systems

by Steffen Hölldobler

INTERNATIONAL CENTER FOR COMPUTATIONAL LOGIC

TECHNISCHE UNIVERSITÄT DRESDEN

Table of Contents

Hölldobler, Kalinke, Störr 1999 [HKS99]

First-order logic programs (acyclic, with injective level mapping) can be approximated by standard sigmoidal feedforward networks.

Existential result via continuity of the T_P-operator.

Features of the approach:

- Shows representability *in principle*.
- Single existing approach for first-order LPs with function symbols.

Limitations of the approach:

- No algorithm how to construct the network.
- Very restricted class of programs.
- Proofs highly dependent on specific assumptions.

Today: Lifting to a more general setting.
Next week: How to construct the networks.

Embedding T_P into the reals

$T_P : I_P \to I_P$, where $I_P = 2^{B_P} \approx 2^{\mathbb{N}}$

$2^{\mathbb{N}}$: all (countably infinite) binary sequences

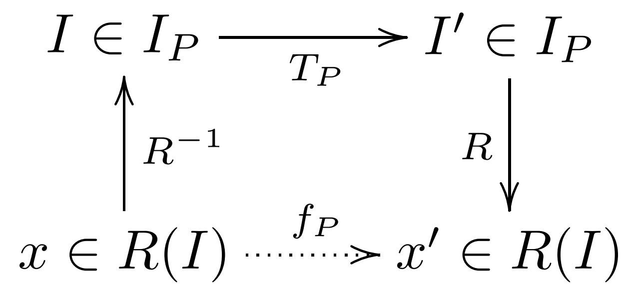
$\rightsquigarrow$ interpret sequence as expansion in b-adic number system! ($b > 2$)

1. Choose an *enumeration* of B_P (= bijective level mapping).
2. Choose a base $b > 2$ for the number system.
3. Choose integers $t \neq f$ with $0 \leq t, f < b$.
4. Given an interpretation $I \subseteq B_P$, set $a_i = f$ if the i-th element of I is false, and set $a_i = t$ otherwise.
5. Set $R(I) = \sum_{i=0}^{\infty} a_i b^i$.

Denote $R(I_P)$ by D_R.

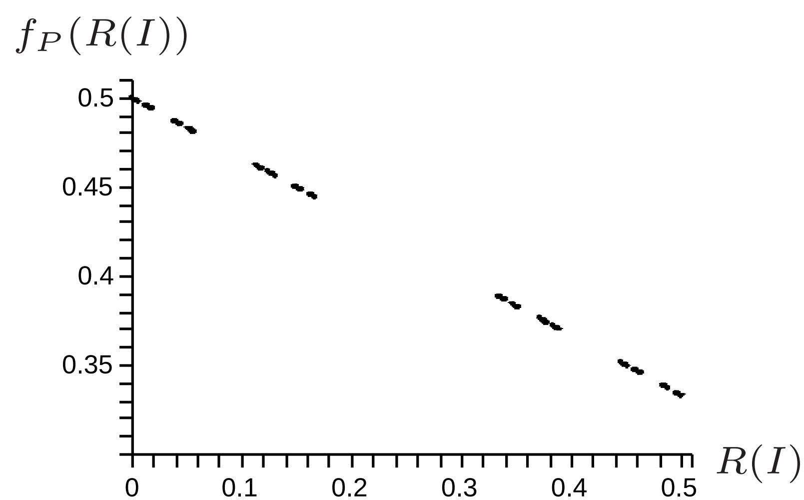
Embedding T_P into the reals

Set $f_P : D_R \to D_R : x \mapsto R(T_P(R^{-1}(x)))$.



even(0).

even$(s(X)) \leftarrow \neg$even(X).



- Domain of f_P is *totally disconnected*. (Is *Cantor set*: discussed later.)
  $\leadsto$ There is a *gap* between any two points.

Approaching Topology

- [Funahashi 89] treats *continuous functions* (on $\mathbb{R}^n$).
- We have seen *metrics* (distance functions) playing a role both on $\mathbb{R}$ and on I_P.
- Iterative behaviour of T_P (*convergence* of iterates) important.

– Continuity, metrics, and convergence are notions known from $\mathbb{R}$.
– They can be studied on much more general spaces (including I_P).
– Corresponding mathematical subject: (Set-theoretic) *Topology*.

Topology:

- Qualitative and quantitative (metric) notions of *distance*.
- Continuity of functions in the sense of *preservation of limits*.
- Notions of approximation in various spaces.
- Abstracts, e.g. allows to link $\mathbb{R}$ and I_P in a sound way.
- Bridge between the continuous (networks) and the discrete (logics).

Table of Contents

Standard (natural) topology on $\mathbb{R}$

Open intervals: known.

Open sets (opens): All unions of open intervals.

Open neighborhood of $x \in \mathbb{R}$: Open set containing x.

The following hold:

(i) All unions of opens are open.

(ii) All *finite* intersections of opens are open.

Conditions (i) and (ii) define: The set of all opens is a *topology* on $\mathbb{R}$.

Note: Infinite intersections of open intervals are not necessarily open.

E.g. $\bigcap_{n=1}^{\infty}]0 - \frac{1}{n}, 1 + \frac{1}{n}[= [0, 1]$.

$\leadsto$ A **topology** $\mathcal{O}$ on a set X is a set of subsets of X s.t. (i) and (ii) hold.

Note: The *empty* union ($=$ all of X) is in every topology!

Standard (natural) topology on $\mathbb{R}$

Standing assumption (for us): There exists a countable subset $\mathcal{B}$ of $\mathcal{O}$ s.t. every open is a union of members of $\mathcal{B}$.

$\mathcal{B}$ called a *base* for $\mathcal{O}$.

Condition called: *$\mathcal{O}$ is second countable (C2)*.

For $\mathbb{R}$: Take as base e.g. all intervals with rational endpoints.

$\mathcal{S}$ is a *subbase* of $\mathcal{O}$ if the set of all finite intersections of sets in $\mathcal{S}$ are a base for $\mathcal{O}$.

Subspace: A subset $X \subseteq \mathbb{R}$ inherits a topology from $\mathbb{R}$, consisting of all $O \cap X$ for all $O \in \mathcal{O}$.

Example: $[0,1]$ as subspace of $\mathbb{R}$.

Some open sets of $[0,1]$: $\emptyset$ $[0,1]$ $[0,0.5[$ $]0.1,0.6[$ ...

Standard (natural) topology on $\mathbb{R}$

Convergence: Sequence (x_n) converges to x (written $x_n \to x$) iff (x_n) is *eventually* in every open neighborhood of x.

Note: Limits are not necessarily unique.
Example: $X = \{0, 1\}$, $\mathcal{O} = \{\emptyset, X\}$.
Then the sequence $1, 1, 1, \ldots$ converges to 0 and 1.

Continuity: A function $f : \mathbb{R} \to \mathbb{R}$ is *continuous* iff the pre-image $f^{-1}(O)$ is open for each open O.

Then: $x_n \to x$ implies $f(x_n) \to x$.
This condition *characterizes* continuity on many spaces (e.g. C2 spaces).

All definitions hold also for more general topologies.

Continuity: Example

Functions on $\mathbb{R}$ are continuous if they *can be drawn without lifting the pen*.

$x \mapsto 2x \qquad x \mapsto x^2 \qquad x \mapsto \ln x \qquad x \mapsto 5\sin\frac{x}{\pi} \;\ldots$

Now consider the subspace $X = [1,2] \cup [3,4] \subseteq \mathbb{R}$.

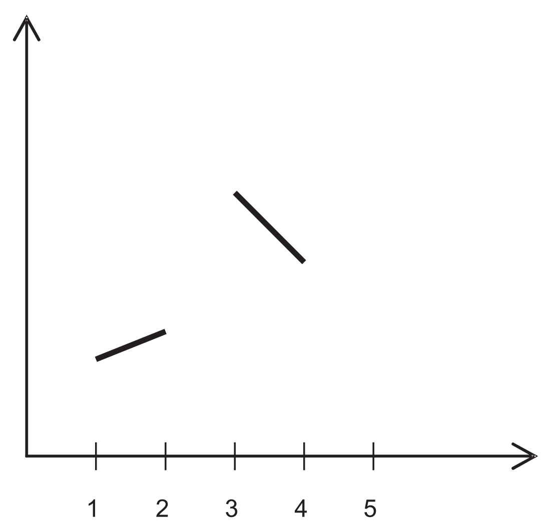
X consists of two connected pieces.

X is *not connected* (as a whole).

$f : X \to \mathbb{R}$

is continuous iff *on each of the connected parts it can be drawn without lifting the pen*.

See example on the right.



Cantor Space

A subspace of $\mathbb{R}$.

$$C = [0,1] \cap \left(\left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right] \right) \cap \left(\left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{1}{3}\right] \cup \left[\frac{2}{3}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right] \right) \cap \ldots$$

- Encoding 0=left, 1=right, each point $x \in C$ is an (infinite) binary sequence $\mathrm{seq}(x)$.
- C is uncountable.
- There is a *gap* between each two points of C.
- A base: $\{B_n(x) \mid x \in C\}$, where $B_n(x) = \{y \in C \mid \mathrm{seq}(x), \mathrm{seq}(y)$ coincide on the first n digits$\}$.

Metrics

A *metric* on a set X is a function $d : X \times X \to \mathbb{R}$ which satisfies the following for all $x, y, z \in X$.

(i) $d(x, y) = 0$ iff $x = y$.

(ii) $d(x, y) = d(y, x)$.

(iii) $d(x, z) \leq d(x, y) + d(y, z)$. (triangle inequality)

- Abstract notion of *distance*.
  E.g. on $\mathbb{R}$: $d(x, y) = |x - y|$.

Collection of all $B_\varepsilon(x) = \{y \in X \mid d(x, y) < \varepsilon\}$ is base of a topology.

But not every topology comes from a metric!

- $B_\varepsilon(x)$ is called the *open ball with center* x *and radius* ε.
- Every metric induces exactly one topology!

Metrics on Cantor Space

Metric inherited from $\mathbb{R}$:

For $x, y \in C$ set $d(x,y) = |x-y|$.

Induced topology: subspace topology inherited from $\mathbb{R}$.

Prefix distance on sequences:

For $x, y \in C$ set $\delta(x,y) = 2^{-n}$,

where n is least s.t. $\mathrm{seq}(x)$ and $\mathrm{seq}(y)$ differ on the n-th digit.

Prefix distance produces base $\{B_{2^{-n}}(x) \mid x \in C\}$, where

$B_{2^{-n}}(x) = \{y \in C \mid \mathrm{seq}(x), \mathrm{seq}(y)$ coincide on the first n digits$\}$.

$\rightsquigarrow$ The base mentioned before!

We will call the prefix distance also *Fitting metric*.

Cantor space

- Can be characterized by the following properties: totally disconnected, compact, Hausdorff, second countable, dense in itself.
- It is a very *well-known* and omnipresent topological space.

- Cantor space sometimes known as *Cantor dust*.
- Cantor space is *self-similar*.

  Zooming in on it produces a very similar picture.
- It is a *fractal* as known from *Chaos theory* or *topological dynamics*.
- Can be "produced" by iterated function systems.

  $\rightsquigarrow$ More about this in two weeks!

Table of Contents

A topology on I_P: The atomic topology Q

[Batarekh & Subrahmanian 1989] [Seda 1995]

For each finite conjunction $C = L_1 \wedge \cdots \wedge L_n$ of literals,
set $\mathcal{G}(C) = \{I \in I_P \mid I \models C\}$.
$\{\mathcal{G}(C) \mid C$ a finite conjunction of literals$\}$ is a *base* of the topology Q.

A sequence (I_n) in I_P converges wrt. Q to some $I \in I_P$ iff

- For each $A \in I$ exists n_0 s.t. for all $k \geq n_0$: $A \in I_k$.
- For each $A \notin I$ exists n_1 s.t. for all $m \geq n_1$: $A \notin I_m$.

A result:

If $T_P^n(K) \to I$ then $I \subseteq T_P(I)$, i.e. I is a *model* of P.
If T_P is also continuous, then $I = T_P(I)$.

A topology on I_P: The atomic topology Q

Another base for Q:

- Fix an enumeration (injective level mapping) on B_P: $A_1, A_2, A_3, \ldots$
- Let C be conjunctions of the form $L_1, \ldots, L_n$, where L_i is A_i or $\neg A_i$.
  Set $\mathcal{G}(C) = \{I \in I_P \mid I \models C\}$ as before.
  Then $\{\mathcal{G}(C) \mid C$ of the indicated form $\}$ is a base of the topology Q.

(Clear: This is a subbase. But it really is also a base!)

Metric characterization of Q:

- $d(I, K) = 2^{-n}$, where n is least such that the enumerations of elements in I and K differ at the n-th position.

$\rightsquigarrow$ Exactly the same as the prefix distance (Fitting metric)!

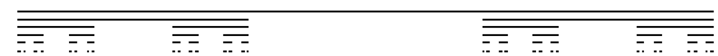
I_P and $\mathbb{R}$ — topological link

Cantor space:

binary sequence

(via the graphical construction)

$\rightsquigarrow x \in C \subseteq \mathbb{R}$



I_P with Q:

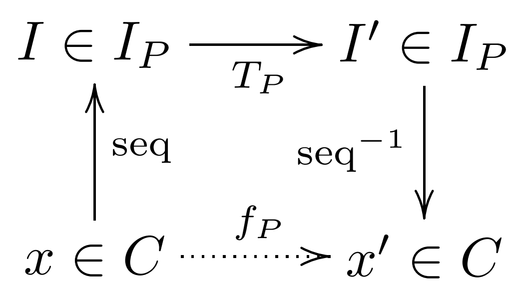
binary sequence (via enumeration of interpretation)

$\rightsquigarrow I \in I_P$

Fitting metric (prefix distance) topologizes both spaces.

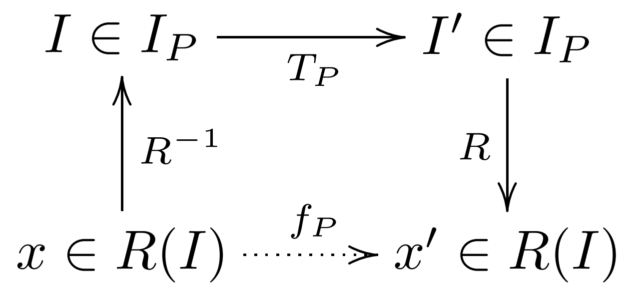
$\Rightarrow I_P$ with Q is topologically *the same* as C!

$\Rightarrow f_P$ is continuous *iff* T_P is!



[HKS99] in topological light

[Hitzler, Hölldobler & Seda, 2004]



- Injective level mapping = enumeration of B_P.
- $R : I_P \to C$ has Cantor space as range.
  (may be different concrete set on $\mathbb{R}$, but topologically the same)
- R is a *topological homeomorphism*
  (mapping preserving the topological structure).
- Acyclicity of the program: guarantees continuity of T_P.
  $\rightsquigarrow$ More about this later.

$\Rightarrow$ Hence: f_P as representation of T_P on $\mathbb{R}$ is continuous and can be approximated by feedforward networks!

Table of Contents

Funahashi's theorem revisited

Theorem (Funahashi 1989, simplified version):

σ sigmoidal
$K \subseteq \mathbb{R}$ compact,
$f : K \to \mathbb{R}$ continuous,
$\varepsilon > 0$.

Then there exists perceptron with sigmoidal σ and I/O-function $\bar{f} : K \to \mathbb{R}$ with

$$\max_{x \in K} \left\{ d \left(f(x), \bar{f}(x) \right) \right\} < \varepsilon;$$

d metric which induces natural topology on $\mathbb{R}$.

I.e. every continuous function $f : K \to \mathbb{R}$ can be uniformly approximated by I/O-functions of perceptrons.

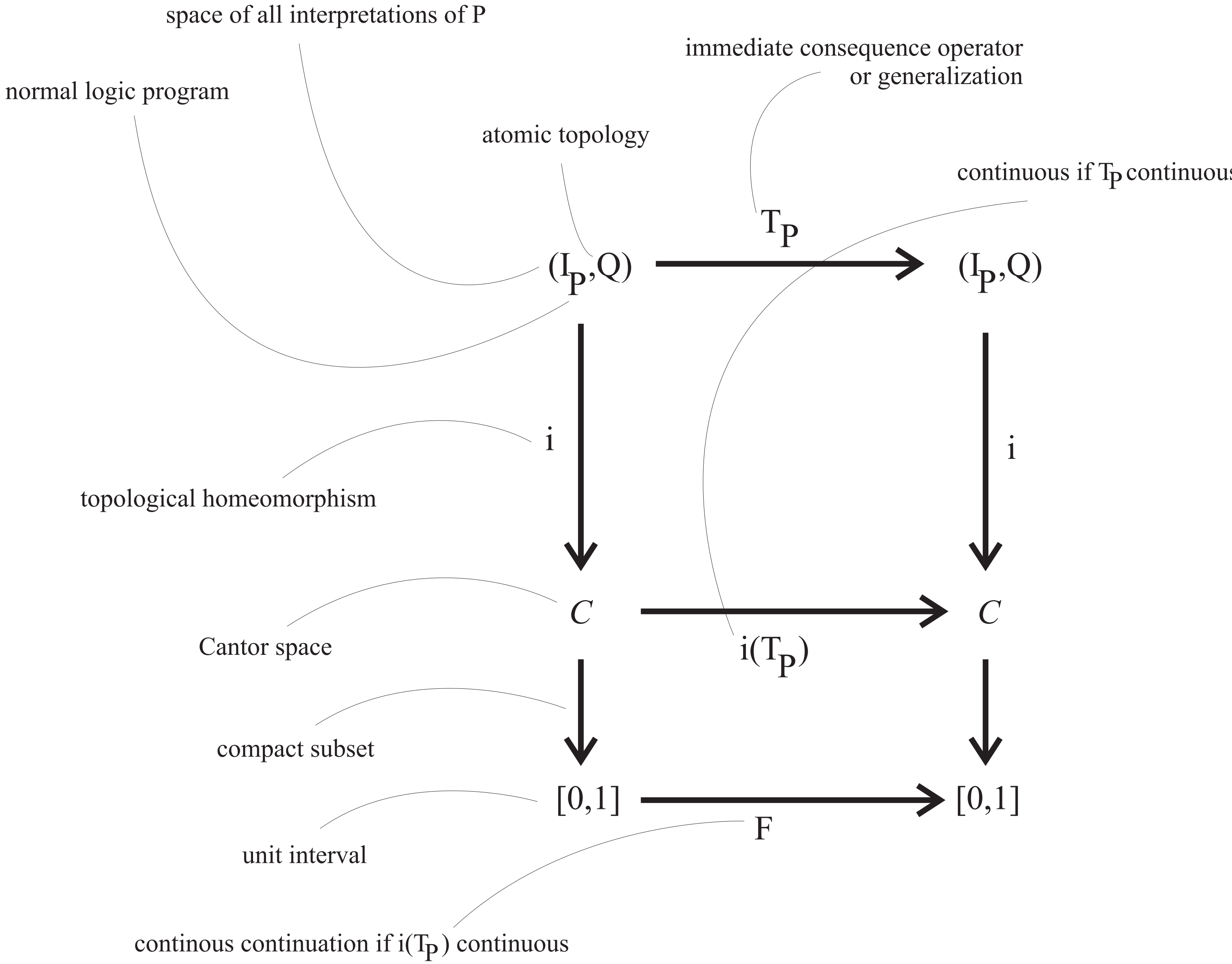

space of all interpretations of P
immediate consequence operator
or generalization
normal logic program
atomic topology
continuous if T_P continuous
T_P
(I_P,Q)
(I_P,Q)
i
i
topological homeomorphism
C
C
Cantor space
$i(T_P)$
compact subset
[0,1]
[0,1]
F
unit interval
continous continuation if $i(T_P)$ continuous

Table of Contents

General consequence operators: multi-valued logics

- Truth values $\mathcal{T} = \{t_1, \dots, t_n\}$.
- Interpretations are functions $I : B_P \to \mathcal{T}$.
- $I_{P,n} = I_P$ set of all interpretations.
- $\mathcal{B}_A$ set of all atoms in bodies of clauses in $\mathrm{ground}(P)$ with head $A \in B_P$.
- $T : I_P \to I_P$ *consequence operator* for P, if for all $I \in I_P$ and all $A \leftarrow \mathtt{body}$ in P we have that $T(I)(A) \leftarrow I(\mathtt{body})$ holds via truth table.
- T *local* if $T(I)(A) = T(K)(A)$ for all $A \in B_P$ and all $I, K \in I_P$ which agree on $\mathcal{B}_A$.
- T_P is a local consequence operator.

Other examples: Operators as defined by Fitting (1985,199x) in three- or four-valued logic.

Cantor topology $\mathcal{Q}$

- For Q, we considered *binary* sequences (two truth values).
- Now consider sequences, where each element can be one out of n values (n truth values).

Use the same prefix distance:

For $I, K \in I_{P,n}$ set $\delta(I, K) = 2^{-n}$,
where n is least s.t. I and K differ on the n-th element.

Topological structure turns out to be the same!

$\Rightarrow$ $\mathcal{Q}$ is *the same as* the Cantor topology!
$\Rightarrow$ $I_{P,n}$ is *the same as* Cantor space!

Approximation of continuous consequence operators

Theorem (Hitzler & Seda 2003)

> Let P be a logic program, T be a locally finite consequence operator, and ι be a homeomorphism from $(I_{P,n}, \mathcal{Q})$ to $\mathcal{C}$.
> Then $\iota(T)$ can be uniformly approximated by I/O-functions of 3lfns.

This holds *mutatis mutandis* e.g. for radial basis function networks (activation function is gaussian).

ι normally given via some enumeration (injective level mapping) $l : B_P \to \mathbb{N}$ and some corresponding p-adic expansion.

There exist *uncountably many homeomorphisms* from I_P to C.
Lots of degrees of freedom!

Characterizing continuity in $\mathcal{Q}$

Consequence operator T on I_P is *locally finite*, if for all $A \in B_P$ and all $I \in I_P$ there exists a finite set $S \subseteq \mathcal{B}_A$ with $T(J)(A) = T(I)(A)$ for all $J \in I_P$ which agree with I on S.

Theorem

A local consequence op. is locally finite iff it is continuous in $\mathcal{Q}$.

Sufficient:

- P is *covered*, i.e. does not contain any *local variables* (occuring in some body, but not in corresponding head).

Continuity of T_P

Continuity of T_P is guaranteed under any of the following conditions:

- P is propositional.
- P does not contain function symbols.
- P is acyclic with respect to an injective level mapping.
- P is covered.

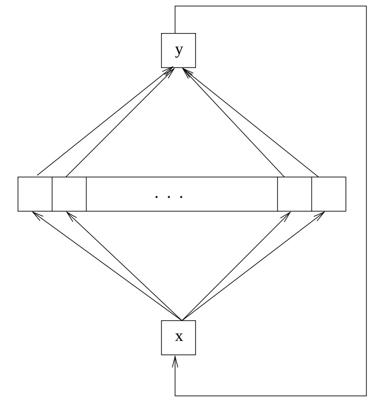
Open Question

Describe a maximal class $\mathcal{A}$ of programs such that for each $P \in \mathcal{A}$ there exists a *covered* program Q with $T_Q = T_P$.

Does $\mathcal{A}$ contain *all* programs P with continuous T_P?
Probably not, but it may contain all
computationally relevant such programs.

Table of Contents

Recursive architecture: Reasoning



T locally finite consequence operator. f approximating I/O-function.

- For all $I \in I_P$ and $n \in \mathbb{N}$ we have $|f^n(\iota(I)) - \iota(T^n(I))| \leq \varepsilon \frac{1-\lambda^n}{1-\lambda}$.

Need: λ Lipschitz-constant of F (i.e. $|F(x) - F(y)| \leq \lambda|x - y|$ for all x, y), and F the continuation of $\iota(T)$ to $[0, 1]$.
ε bound on approximation error.

- If F is contraction, then $\left(F^k(\iota(I))\right)$ converges for all I to unique fixed point x of F and $\exists m \in \mathbb{N}\ \forall n \geq m$: $|f^n(\iota(I)) - x| \leq \varepsilon \frac{1}{1-\lambda}$.
Furthermore, T is a contraction on the complete space I_P (with suitable metric), and we have $\iota(M) = x$ for the unique fixed point M of T.

- Assume there is $I \in I_P$ s.t. $T^n(I)$ converges in $\mathcal{Q}$ to a fixed point M of T.

Then for every $\delta > 0$ there exists some $n \in \mathbb{N}$ and a network with $|f^n(\iota(I)) - \iota(M)| < \delta$.

Table of Contents

Remarks on measurability

An alternative to [Funahashi 89]:

Theorem (Hornik, Stinchcombe, White 1989, simplified version)

$\sigma : \mathbb{R} \to (0,1)$ monotonic increasing, onto.
$f : \mathbb{R} \to \mathbb{R}$ Borel measurable,
μ Borel probability measure on $\mathbb{R}$,
$\varepsilon > 0$.

Then there is a perceptron with sigmoidal activation function σ and I/O-function $\bar{f} : \mathbb{R} \to \mathbb{R}$ with

$$\varrho_\mu(f, \bar{f}) = \inf \left\{ \delta > 0 : \mu \left\{ x : \left| f(x) - \bar{f}(x) \right| > \delta \right\} < \delta \right\} < \varepsilon.$$

I.e. the set of I/O-functions which can be computed using 3lfns is dense with respect to ϱ_μ in the set of all Borel measurable functions $f : \mathbb{R} \to \mathbb{R}$.

Measurable consequence operators

Theorem (Hitzler & Seda 2000)

> Local consequence operators are always measurable with respect to $\sigma(\mathcal{Q})$.

But:

Approximation by networks is only *almost everywhere*.

Cantor set has measure 0.

Result can be improved to some extent, but principal problem remains.
$\Rightarrow$ Continuity approach appears to be more promising!

Table of Contents

Non-monotonic reasoning

Fixed points of operator GL_P yield the *stable models* of P.
(as in Answer Set Programming)

[DK89] Phan M. **Dung** and Kanchana **Kanchanasut**, A fixpoint approach to declarative semantics of logic programs. Proc. NACLP'89, 1989.

Program transformation $P \mapsto \mathrm{fix}(P)$.
Complete unfolding through positive body literals.

[Wen02] Matthias **Wendt**, Unfolding the well-founded semantics, Journal of Electrical Engineering 2002.

Shows $\mathsf{GL}_P(I) = T_{\mathrm{fix}(P)}(I)$ for all interpretations I.

$\rightsquigarrow$ Allows to carry over results. (Bader & Hitzler, in preparation)
$\rightsquigarrow$ Works similarly also for well-founded semantics.

The Fixpoint Completion

Quasi-interpretation K: set of clauses of form $A \leftarrow \neg B_1, \dots, \neg B_m$.

Program P: set of (ground) clauses of form

$$A \leftarrow A_1, \dots, A_n, \neg B_1, \dots, \neg B_m.$$

$T'_P(K)$ set of $A \leftarrow \texttt{body}_1, \dots, \texttt{body}_n, \neg B_1, \dots, \neg B_m$
where $A \leftarrow A_1, \dots, A_n, \neg B_1, \dots, \neg B_m$ in P
and $A_i \leftarrow \texttt{body}_i$ in K for all i.

$T'_P \uparrow \omega = \mathrm{fix}(T'_P) = \mathrm{fix}(P)$ quasi-interpretation.

Thank You!

Pascal Hitzler — Universität Karlsruhe (TH)

Sebastian Bader, Steffen Hölldobler — International Center for Computational Logic, Technische Universität Dresden

www.neural-symbolic.org