Approximating first-order logic programs by feedforward networks

CLCS SS 2004 — Subsitute Lecturing

Dr. Pascal Hitzler

Artificial Intelligence Institute, Department of Computer Science
Dresden University of Technology, Dresden, Germany
phitzler@inf.tu-dresden.de, www.wv.inf.tu-dresden.de/∼pascal/

July 14, 2004

Abstract

We want to apply Funahashi's theorem [Fun89] in order to approximate the T_P operator for first-order (normal) logic programs P via 3-layer feedforward networks. I.e. we need to understand T_P as a continuous function on the reals.

We will need to study some preliminaries from *set-theoretic topology* first — main reference is [Wil70]. We will then work towards a very recent research result from [HS00, HHS0x], which extends results from [HKS99]. We close with some further considerations about the methods and results.

Some exercises are of central importance for subsequent material. They are marked with a $(*)$.

Contents

1 Topologies

In the following, let X be a set. By $\mathcal{P}(X)$ we denote the powerset of X. Until further notice, B_P will denote the set of all atoms of some first-order language — we think of it as the first-order language underlying some logic program P. The set $I_P = \mathcal{P}(B_P)$ is the corresponding set of all interpretations.

1.1 Open sets, closed sets, neighborhoods

1.1 Definition A *topology* $\mathcal{O}$ on X is a subset of $\mathcal{P}(X)$ with the following properties:

(T1) $\emptyset, X \in \mathcal{O}$

(T2) If I is a set and $O_i \in \mathcal{O}$ for each $i \in I$, then $\bigcup_{i\in I} O_i \in \mathcal{O}$.item If $O_1, O_2 \in \mathcal{O}$, then $O_1 \cap O_2 \in \mathcal{O}$.

The pair $(X, \mathcal{O})$ (or simply X) is called a *topological space*. Each $O \in \mathcal{O}$ is called *open* in $\mathcal{O}$. A set $A \subseteq X$ is called *closed* in $\mathcal{O}$ if its complement ${}^cA$ is open in $\mathcal{O}$. A set $U \subseteq X$ is called a *neighborhood* of a point $x \in X$ if $x \in O \subseteq U$ for some $O \in \mathcal{O}$.

1.2 Example $(X, \mathcal{P}(X))$ and $(X, \{\emptyset, X\})$ are topologies on X, called the *discrete* and the *indiscrete* topology, respectively.

Exercise 1 ($*$)
Determine all topologies on the two-point set $\{0, 1\}$.

Exercise 2
Show that a set is open if and only if it contains a neighborhood for each of its points.

1.2 Bases and subbases

1.3 Definition Let $(X, \mathcal{O})$ be a topological space. A *base* of $\mathcal{O}$ is a subset $\mathcal{B} \subseteq \mathcal{O}$ such that each $O \in \mathcal{O}$ is the union of members of $\mathcal{B}$. A *subbase* of $\mathcal{O}$ is a subset $\mathcal{S} \subseteq \mathcal{O}$ such that the set of all finite intersections of members of $\mathcal{S}$ is a base for $\mathcal{O}$, or equivalently, if each member of $\mathcal{O}$ is the union of finite intersections of members of $\mathcal{S}$.

1.4 Proposition Let X be a set. $\mathcal{B} \subseteq \mathcal{P}(X)$ is a base of a topology if and only if $\bigcup \mathcal{B} = X$ and for all $U, V \in \mathcal{B}$ and each $x \in U \cap V$ there exists $W \in \mathcal{B}$ with $x \in W$ and $W \subseteq U \cap V$. Furthermore, if $\mathcal{B}$ is a base of $\mathcal{O}$, then $\mathcal{O}$ is the smallest topology containing each set in $\mathcal{B}$.

Proof: If $\mathcal{B}$ is a base for $\mathcal{O}$, $U, V \in \mathcal{B}$ and $x \in U \cap V \in \mathcal{O}$, then $U \cap V$ is the union of members of $\mathcal{B}$ and hence there exists $W \in \mathcal{B}$ with $x \in W \subseteq U \cap V$. Conversely, let $\mathcal{B}$ be a family with the specified property and let $\mathcal{O}$ be the family of all unions of members of $\mathcal{B}$. We obtain $\bigcup \mathcal{B} = X \in \mathcal{B}$ and $\bigcup \emptyset = \emptyset \in \mathcal{B}$ which shows that (T1) holds. To see (T2), note that a union of members of $\mathcal{O}$ is itself a union of members of $\mathcal{B}$. For (T3), Let $U, V \in \mathcal{O}$ and $x \in U \cap V$. Choose $U', V' \in \mathcal{B}$ such that $x \in U' \subseteq U$

and $x \in V' \subseteq V$, and then $W \in \mathcal{B}$ with $x \in W \subseteq U' \cap V' \subseteq U \cap V$. So $U \cap V$ can be expressed as a union of members of $\mathcal{B}$. To prove the last statement, note that each topology containing $\mathcal{B}$ must contain all unions of members of $\mathcal{B}$, i.e. must contain $\mathcal{O}$. Since $\mathcal{O}$ is itself a topology, it is the smallest topology containing $\mathcal{B}$. ■

1.5 Proposition Let X be a set. Every $\emptyset \neq \mathcal{S} \subseteq \mathcal{P}(X)$ is the subbase of a topology. Furthermore, if $\mathcal{O}$ is a topology with subbase $\mathcal{S}$ then $\mathcal{O}$ is the smallest topology which contains every set in $\mathcal{S}$.

Proof: Let a set $\mathcal{S}$ be given and let $\mathcal{B}$ be the set of all finite intersections of member of $\mathcal{S}$. Then $X = \bigcap \emptyset \in \mathcal{B}$ and the intersection of two members of $\mathcal{B}$ is again a member of $\mathcal{B}$ and Proposition 1.4 yields that $\mathcal{B}$ is the base of a topology. Let $\mathcal{B}$ be the base of a topology $\mathcal{O}$. Since every base of a topology which contains $\mathcal{S}$ must contain $\mathcal{B}$, it must also contain $\mathcal{O}$. Since $\mathcal{O}$ is itself a topology, it is the smallest topology containing $\mathcal{S}$. ■

1.6 Example (a) The set of all open intervals of $\mathbb{R}$ is the base for a topology on $\mathbb{R}$, called the *natural topology* on $\mathbb{R}$.

(b) Let (X, d) be a metric space. A set of the form $B_\varepsilon(x) = \{y \in X : d(x, y) < x\}$ is called an (*open*) *ball* in X. The set of open balls forms the base for a topology. This topology is called the *topology induced by* d and is denoted by $\mathcal{O}_d$.

Exercise 3 (∗)
Show that the statements from Example 1.6 are correct.

1.7 Example ([Sed95]) Consider the space of interpretations $I_P = \mathcal{P}(B_P)$. For $A \in B_P$ set $\mathcal{G}(A) = \{I \in I_P : A \in I\}$ and $\mathcal{G}(\neg A) = \{I \in I_P : A \notin I\}$. The topology with subbase $\{\mathcal{G}(A) : A \in B_P\}$ is called the *Scott topology* on I_P. The topology with subbase $\{\mathcal{G}(A) : A \in B_P\} \cup \{\mathcal{G}(\neg A) : A \in B_P\}$ is called the *atomic topology* Q on I_P.

1.8 Proposition Show that the sets of the form $\mathcal{G}(L_1 \wedge \cdots \wedge L_n) = \{I \in I_P : I \models L_1 \wedge \cdots \wedge L_n\}$, where $L_1, \ldots, L_n$ are literals, form a base for the atomic topology on I_P.

Proof: Obviously, $\{I \in I_P : I \models L_1 \wedge \cdots \wedge L_n\} = \bigcap_{i=1}^{n} \mathcal{G}(L_i)$, so $\mathcal{G}(L_1 \wedge \cdots \wedge L_n) \in Q$, so the result follows from Proposition 1.4. ■

Exercise 4
The natural metric *on* $\mathbb{R}$ *is defined by* $d(x, y) = |x - y|$.

(a) Show that the natural metric induces the natural topology on $\mathbb{R}$.

(b) Show that the set of all balls of the form $B_\varepsilon(x)$*, where* $\varepsilon, x \in \mathbb{Q}$*, forms a base for the natural topology on* $\mathbb{R}$.

1.3 Subspaces and product spaces

1.9 Definition Let $Y \subseteq X$. The *subspace topology* on Y (induced by the topology $\mathcal{O}$ on X) consists of all sets of the form $O \cap Y$, where $O \in \mathcal{O}$.

Exercise 5
Show that the subspace topology is indeed a topology.

Exercise 6
Let (X, d) be a metric space and $Y \subseteq X$. Define $d' : Y \times Y \to \mathbb{R}$ by setting $d'(x, y) = d(x, y)$.

(a) Show that d' is a metric.

(b) Show that the topology induced by d' coincides with the subspace topology on Y induced by the topology $\mathcal{O}_d$ on X.

Cantor set $\mathcal{C}$ is a subset on the real line which consists of all real numbers which can be written in the form

$$\sum_{i=1}^{\infty} a_i 3^{-i},$$

where $a_i \in \{0, 2\}$ for all i. The representation of each $a \in \mathcal{C}$ as such a power series is unique (why?), and we can hence identify each $a \in \mathcal{C}$ uniquely with a sequence $(a_i)_{i \in \mathbb{N}}$. Now define $d : \mathcal{C} \times \mathcal{C} \to \mathbb{R}$ by setting $d((a_i)_i, (b_i)_i) = 2^{-n}$, where n is the smallest natural number such that $a_n \neq b_n$.

Exercise 7 (∗)
Show that the function d just defined is an ultra*metric on $\mathcal{C}$, i.e. that it is a metric, and furthermore for all $x, y, z \in \mathcal{C}$ we have $d(x, z) \leq \max\{d(x, y), d(y, z)\}$.*

1.10 Proposition The topology induced by d coincides with the subspace topology which $\mathcal{C}$ inherits from $\mathbb{R}$.

Proof: Recall that for $|r| < 1$ we have $\sum_{i=1}^{n} r^i = \frac{r(1-r^n)}{1-r}$ and $\sum_{i=1}^{\infty} r^i = \frac{r}{1-r}$.

We first show that $d(x, y) < 2^{-n}$ implies $|x - y| < 3^{-(n-1)}$. Indeed, let $d(x, y) \leq 2^{-(n+1)}$. Then (x_i) and (y_i) agree up to the n-th element and differ on the $(n + 1)$-st element. But this means that

$$\begin{aligned}
|x - y| &= \left| \sum_{i=1}^{\infty} x_i 3^{-i} - \sum_{i=1}^{\infty} y_i 3^{-i} \right| = \left| \sum_{i=n+1}^{\infty} \left(x_i 3^{-i} - y_i 3^{-i} \right) \right| \leq \sum_{i=n+1}^{\infty} \left| x_i 3^{-i} - y_i 3^{-i} \right| \\
&= \sum_{i=n+1}^{\infty} |x_i - y_i|\, 3^{-i} \leq \sum_{i=n+1}^{\infty} 2 \cdot 3^{-i} = 2 \cdot \sum_{i=n+1}^{\infty} 3^{-i} = 2 \cdot \left(\frac{1}{2} - \left(\frac{1}{2} - \frac{1}{2 \cdot 3^n} \right) \right) \\
&= 3^{-n} < 3^{-(n-1)}.
\end{aligned}$$

Now let $O = \{y : |x - y| < \varepsilon\}$ be a basic open set in the subspace topology on $\mathcal{C}$. Then for each $z \in O$ there exists n such that $U_z = \{y : |z - y| < 3^{n-1}\} \subseteq O$, and by the assertion just shown and by setting $O_z = \{y : d(x, y) < 2^{-n}\}$ we obtain

$z \in O_z \subseteq U_z \subseteq O$. Hence $O = \bigcup_{z \in O} O_z$ which shows that O is open in the topology induced by d.

Conversely — by contraposition —, if $d(x,y) \geq 2^{-n}$, then for the least $k \in \mathbb{N}$ for which (x_i) and (y_i) differ on the k-th element, we have $k \leq n$. Consequently,

$$\begin{aligned} |x-y| &= \left| \sum_{i=1}^{\infty} (x_i - y_i)\, 3^{-i} \right| = \left| \sum_{i=k}^{\infty} (x_i - y_i)\, 3^{-i} \right| \\ &= \left| (x_k - y_k) 3^{-k} + \sum_{i=k+1}^{\infty} (x_i - y_i)\, 3^{-i} \right| . \end{aligned}$$

By symmetry, we can assume without loss of generality that $x_i = 2$ and $y_i = 0$, so that

$$\begin{aligned} |x-y| \geq \left| 2 \cdot 3^{-k} - \left| \sum_{i=k+1}^{\infty} (x_i - y_i)\, 3^{-i} \right| \right| &\geq \left| 2 \cdot 3^{-k} - \sum_{i=k+1}^{\infty} 2 \cdot 3^{-i} \right| \\ = 2 \cdot 3^{-k} - 3^{-k} = 3^{-k} \geq 3^{-n}. \end{aligned}$$

We have just shown by contraposition that $|x-y| < 3^{-n}$ implies $d(x,y) < 2^{-n}$.

Now let $O = B_{2^{-n}}(x)$ be a basic open set with respect to d, and let $z \in O$ and $U_z = \{y : d(z,y) < 2^{-n}\}$. Since d is an ultrametric we obtain $d(y,x) \leq \max\{d(y,z), d(z,x)\} < 2^{-n}$ for all $y \in U_z$, and hence $U_z \subseteq O$. By the assertion shown in the previous paragraph we get $O_z \subseteq U_z$ for $O_z = \{y : |y-z| < 3^{-n}\}$, and hence $O = \bigcup_{z \in O} O_z$ which shows that O is open with respect to the subspace topology which $\mathcal{C}$ inherits from $\mathbb{R}$. ■

Exercise 8 (∗)
Show that the collection of sets of the form $\{y \in \mathcal{C} : d(y,x) < 2^{-n}\}$ *is a base of* $(\mathcal{C}, d)$*, and the collection of sets of the form* $\{y \in \mathcal{C} : |x-y| < 3^{-n}\}$ *is a base of* $(\mathcal{C}, |\cdot|)$.

1.11 Definition Cantor set endowed with the subspace topology inherited from $\mathbb{R}$ is called *Cantor space*.

Given a topological space $(X, \mathcal{O})$ and a set Y, let X^Y denote the set of all functions $f : Y \to X$. For $y \in Y$ we write f_y for $f(y)$, and we can identify functions $f \in X^Y$ with families $(f_y)_{y \in Y}$ of elements of X.

1.12 Definition A base for the *product topology* on X^Y is given as follows. The sets of the base are exactly the sets of the form

$$\left\{ f \in X^Y : f_y \in O_y \right\},$$

where $O_y \in \mathcal{O}$ for all $y \in Y$ and $O_y = X$ for all but finitely many y.

1.13 Proposition Let $\mathbf{2}$ be the set $\{0,1\}$ endowed with the discrete topology. Then the product topology on $\mathbf{2}^{B_P}$ coincides with the atomic topology on I_P.

Proof: We need to identify $\mathbf{2}^{B_P}$ and I_P first. This is done by the usual identification of $f : B_P \to \mathbf{2}$ with $\{A \in B_P : f(A) = 1\}$, or conversely, by identifying each $\{A_1, \dots, A_n\} \in \mathcal{P}(B_P)$ with the function f with $f(A_1) = \dots = f(A_n) = 1$ and $f(B) = 0$ for all other B. Each basic open set

$$\mathcal{G}(A_1 \wedge \dots \wedge A_n \wedge \neg B_1 \wedge \dots \wedge B_m) = \{I \in I_P \mid I \models A_1 \wedge \dots \wedge A_n \wedge \neg B_1 \wedge \dots \wedge B_m\},$$

where the A_i and B_j are atoms, thus coincides exactly with the basic open set

$$\{f \in \mathbf{2}^{B_P} : f(A_i) = 1 \text{ for all } i \text{ and } f(B_j) = 0 \text{ for all } j\}.$$

■

Exercise 9
Consider the topological space $\mathcal{S} = (\{0,1\}, \{\emptyset, \{1\}, \{0,1\}\})$, *called* Sierpinski space. *Show that the product topology on* $\mathcal{S}^{B_P}$ *coincides with the Scott topology on* I_P.

1.4 Convergence and continuity

1.14 Definition We say that a sequence $(x_n)_{n\in\mathbb{N}}$ in a topological space $(X, \mathcal{O})$ *converges to* some $x \in X$, written $\lim x_n = x$ or $x_n \to x$, if for each open neighborhood O of x there exists some $n_0 \in \mathbb{N}$ such that $x_m \in O$ for all $m \geq n$. We call x a *limit* of $(x_n)_{n\in\mathbb{N}}$.

Exercise 10
Show that in Definition 1.14 it suffices to consider basic *or* subbasic *open neighborhoods (i.e. open neighborhoods which belong to some fixed base or subbase of the topology* $\mathcal{O}$*).*

Exercise 11
Show that a sequence converges with respect to a metric if and only if it converges with respect to the topology induced by the metric.

1.15 Proposition Sequences in metric spaces have at most one limit. This is not true for all topological spaces.

Proof: Assume $x_n \to x$ and $x_n \to y$. Then $d(x,y) \leq d(x,x_n) + d(x_n,y)$ for all n, and since $d(x,x_n) + d(x_n,y) \to 0$ we have $d(x,y) = 0$, hende $x = y$.

For the second part, consider the indiscrete topology on $\{0,1\}$. Then the constant sequence $0, 0, \dots$ converges both to 0 and to 1. ■

1.16 Proposition Consider I_P endowed with the atomic topology Q. Then a sequence $(I_n)_{n\in\mathbb{N}}$ in I_P converges in Q to some $I \in I_P$ if and only if the following two conditions are satisfied.

(1) For each $A \in I$ there exists $n_1 \in \mathbb{N}$ such that for all $m \geq n_1$ we have $A \in I_m$.

(2) For each $A \notin I$ there exists $n_2 \in \mathbb{N}$ such that for all $k \geq n_2$ we have $A \notin I_k$.

Proof: First assume $I_n \to I$ in Q. Let $A \in B_P$ be arbitrarily chosen. If $A \in I$ then $I \in \mathcal{G}(A)$, hence there exists n_1 such that $I_m \in \mathcal{G}(A)$ for all $m \geq n_1$, which suffices. If $A \notin I$ then $I \in \mathcal{G}(\neg A)$, hence there exists n_2 such that $I_k \in \mathcal{G}(\neg A)$ for all $k \geq n_2$, which suffices.

Conversely, assume that conditions (1) and (2) hold relative to (I_n) and I. Let U be a basic open neighborhood of I, and of the form $\mathcal{G}(A_1 \wedge \cdots \wedge A_n \wedge \neg B_1 \wedge \cdots \wedge B_m)$, which implies for all i and j that $A_i \in I$ i and $B_j \notin I$. By assumption, we can choose n_0 such that for all I_n with $n \geq n_0$ and for all i and j that $A_i \in I_n$ i and $B_j \notin (I_n)$, hence $I_n \in U$ for all $n \geq n_0$ as required. ■

Exercise 12
Consider I_P endowed with the Scott topology. Show that any sequence $(I_n)_{n\in\mathbb{N}}$ in I_P converges, and that the set of its limits achieves a greatest element. Furthermore, if there exists some $n_0 \in \mathbb{N}$ such that the subsequence $(I_n)_{n\geq n_0}$ is monotonically increasing, then the greatest limit of the sequence $(I_n)_{n\in\mathbb{N}}$ is $\sup\{I_n : n \geq n_0\}$.

1.17 Definition Let $(X, \mathcal{O})$ and $(Y, \mathcal{O}')$ be topological spaces. A function $f : X \to Y$ is called *continuous at* $x \in X$ if $f^{-1}(O) \in \mathcal{O}$ for each open neighborhood $O \in \mathcal{O}'$ of $f(x)$. A function is called *continuous* if it is continuous at all points of its domain.

1.18 Proposition Let f be continuous and assume $x_n \to x$. Then $f(x_n) \to f(x)$.

Proof: Let U be some open neighborhood of $f(x)$. Then $f^{-1}(U)$ is an open neighborhood of x. Hence there exists n_0 such that $x_n \in U$ for all $n \geq n_0$. Consequently, $f(x_n) \in U$ for all $n \geq n_0$. ■

Exercise 13
Show that it suffices to consider basic or subbasic open neighborhoods in Definition 1.17.

Let (X, d) and (Y, d') be metric spaces. A function $f : X \to Y$ is (metrically) *continuous* (with respect to d and d') if for each $x \in X$ and for all $\varepsilon > 0$ there exists $\delta > 0$ such that for all y with $d(x, y) < \delta$ we have $d(f(x), f(y)) < \varepsilon$.

Exercise 14
Show that a function is metrically continuous if and only if it is continuous with respect to the topologies induced by the metrics.

Exercise 15
Show that a function on I_P is continuous with respect to the Scott topology if and only if for any monotonically increasing sequence $(I_n)_{n\in\mathbb{N}}$ we have $f(\sup I_n) = \sup f(I_n)$.

2 Representing first-order logic programs by artificial neural networks

2.1 Applying Funahashi's Theorem

We assume for the following, that B_P is (countably) infinite.

2.1 Definition Let X and Y be topological spaces. A function $\iota : X \to Y$ is called a *homeomorphism* if it is a continuous bijection and ι^{-1} is also continuous. In this case, we call X and Y *homeomorphic.*

2.2 Theorem The following spaces are homeomorphic.

(i) $(\mathcal{C}, \mathcal{O}_{|\cdot|})$.

(ii) $(\mathcal{C}, \mathcal{O}_d)$, where d is as in Proposition 1.10.

(iii) (I_P, Q).

(iv) $\mathbf{2}^{B_P}$, where $\mathbf{2}$ carries the discrete topology.

Proof: It follows from Proposition 1.10, that (i) and (ii) are homeomorphic — the identity function acts as homeomorphism. Likewise, it follows from Proposition 1.13 that (iii) and (iv) are homeomorphic, and the bijection $f \mapsto \{A \in B_P : f(A) = 1\}$ acts as homeomorphism.

We show that (ii) and (iii) are homeomorphic, which finishes the proof. First choose an arbitrary bijective level mapping $l : B_P \to \mathbb{N}$, and define

$$R : I_P \to \mathcal{C} : I \mapsto \sum_{i=1}^{\infty} g_I\left(l^{-1}(i)\right) 3^{-i},$$

where $g_I(A) = 2$ if $A \in I$ and $g_I(A) = 0$ if $A \notin I$. Obviously, R is a bijection between I_P and $\mathcal{C}$. We show that it is a homeomorphism.

Let $B = B_{2^{-n}}(x)$ be an arbitrary basic open ball with respect to d, and let $x = (x_i)_{i\in\mathbb{N}}$ with $x_i \in \{0, 2\}$. Then for all $y = (y_i)_{i\in\mathbb{N}} \in B$ we have that $x_i = y_i$ for all $i \leq n$. Now consider the open set $G = \mathcal{G}(h_1(x_1) \wedge \cdots \wedge h_n(x_n))$, where $h_i(0) = \neg l^{-1}(i)$ and $h_i(2) = l^{-1}(i)$ for all i. We obtain $R^{-1}(B) = G$. Since B was arbitrarily chosen, we have shown that R is continuous.

In order to show that R^{-1} is continuous, let $G = \mathcal{G}(L)$ be an arbitrary subbasic open set with respect to Q and let $l(L) = k$. Assume first that $L = A$ is an atom. Let K be the set of all elements of $\mathcal{C}$ of the form $(x_i)_{i\in\mathbb{N}}$ with $x_k = 2$. Consider $U = \bigcup_{x\in K} \{B_{2^{-(k+1)}}(x)\}$, which is open with respect to d, and $R(G) = U$. If L is a negated atom, then the argument is similar. This completes the proof. ∎

Exercise 16
Complete the proof of Theorem 2.2 by spelling out the argument for the case when L is a negated atom.

For sets X, Y, and a function $f : X \to X$ and a bijection $g : X \to Y$, define $g(f) : Y \to Y : y \mapsto g(f(g^{-1}(y)))$.

Exercise 17
Let $(X, \mathcal{O})$ and $(X', \mathcal{O}')$ be topological spaces and $\iota : X \to X'$ be a homeomorphism. Show that $f : X \to X$ is continuous if and only if $\iota(f)$ is continuous.

2.3 Theorem (Funahashi) Suppose that $\phi : \mathbb{R} \to \mathbb{R}$ is non-constant, bounded, monotone increasing and continuous. Let $K \subseteq \mathbb{R}^n$ be compact, let $f : K \to \mathbb{R}$ be a continuous mapping and let $\varepsilon > 0$. Then there exists a 3-layer feedforward network with activation function ϕ whose input-output mapping $\bar{f} : K \to \mathbb{R}$ satisfies $\max_{x \in K} d(f(x), \bar{f}(x)) < \varepsilon$, where d is a metric which induces the natural topology on $\mathbb{R}$.

2.4 Theorem (Main Result) Let P be a program such that T_P is continuous in Q, and let ι be a homeomorphism from $(I_{P,n}, \mathcal{Q})$ to $\mathcal{C}$. Then T_P (more precisely $\iota(T_P)$) can be uniformly approximated (in the sense of Funahashi's theorem) by input-output mappings of 3-layer feedforward networks.

Proof: Since T_P is continuous in Q, and ι is a homeomorphism, we obtain that $\iota(T_P)$ is continuous as a function in Cantor space $\mathcal{C}$. Since $\mathcal{C}$ is a compact subspace of $\mathbb{R}$, the result follows by applying Funahashi's theorem. ■

2.5 Research Problem Theorem 2.4 is not constructive, i.e. we currently do not know how to obtain approximating networks from given programs. Ideally, we would like to be able to read off network parameters directly from the program.

2.6 Research Problem How to use Theorem 2.4 as the base for an integrated neural-symbolic learning system?

Exercise 18
Show that $\mathcal{C}$ is compact as a subset of $\mathbb{R}$.

The choice of homeomorphism using the ternary number system is very special. Indeed, there exist uncountably many homemomorphisms of Cantor space with itself.

Also, the representation of Cantor space in the real line which we used is very special. There exist many other ways of characterizing Cantor space as a subset of the reals.

2.7 Theorem ([Wil70]) The second countable totally disconnected compact Hausdorff spaces which are dense in itself are exactly those topological spaces which are homeomorphic to Cantor space.

A space X is *second countable* iff it has a countable base. It is *Hausdorff* (or T_2) iff for all $x \neq y$ there exist open neighborhoods O of x and U of y such that $O \cap U = \emptyset$. It is *compact* iff for each collection $(O_i)_{i \in I}$ of open sets with $X \subseteq \bigcup_{i \in I} O_i$ there exists a finite subset $K \subseteq I$ such that $X \subseteq \bigcup_{k \in K} O_k$. It is *dense in itself* (sometimes called *perfect*, or *without isolated points*), iff no singleton set (containing exactly one element) is open. It is *totally disconnected* iff for each $x \neq y$ there exist disjoint open neighborhoods U of x and V of y such that $U \cup V = X$.

Exercise 19
Show that Cantor space is a second countable totally disconnected Hausdorff space which is dense in itself.

2.2 T_P and the Cantor topology

2.8 Definition For $A \in B_P$ let $\mathcal{B}_A$ denote the set of all atoms which occur in bodies of ground instances of clauses with head A. We call T_P *locally finite for $A \in B_P$ and $I \in I_P$* if there exists a finite subset $S = S(A, I) \subseteq \mathcal{B}_A$ such that for al $J \in I_P$ which agree with I on S we have $A \in T_P(J)$ iff $A \in T_P(I)$. We say that T_P is *locally finite* if it is locally finite for all $A \in B_P$ and all $I \in I_P$.

Exercise 20
Show the following.

(a) *If $A \in T_P(I)$, then T_P is locally finite for A and I.*

(b) *T_P is continuous in Q if and only if it is locally finite for all A and I with $A \notin T_P(I)$.*

2.9 Theorem T_P is continuous in Q if and only if it is locally finite.

Proof: For $A \in B_P$, let $\mathcal{B}_A$ denote the set of all atoms which occur in bodies of ground instances of clauses of P with head A.

Assume first that T_P is locally finite. Let $I \in I_P$, let $A \in B_P$, and let $G_2 = \mathcal{G}(A)$ be a subbasic neighbourhood of $T_P(I)$ in Q. Since T_P is locally finite, there is a finite set $S \subseteq \mathcal{B}_A$ such that $T_P(J)(A) = T_P(I)(A)$ for all $J \in \bigcap_{B \in S} \mathcal{G}(B)$. By finiteness of S, the set $\bigcap_{B \in S} \mathcal{G}(B)$ is open and contains I, and this suffices for continuity of T_P.

For the converse, assume that T_P is continuous in Q and let $A \in B_P$ and $I \in I_P$ be chosen arbitrarily. Then $G_2 = \mathcal{G}(A)$ is a subbasic open set, so that, by continuity of T_P, there exists a basic open set $G_1 = \mathcal{G}(B_1) \cap \cdots \cap \mathcal{G}(B_k)$ with $T_P(G_1) \subseteq G_2$. In other words, we have $T_P(J)(A) = T_P(I)(A)$ for each $J \in \bigcap_{B \in S'} \mathcal{G}(B)$, where $S' = \{B_1, \ldots, B_k\}$ is a finite set. Note that the value of $T_P(J)(A)$ depends only on the values $J(A)$ of atoms $A \in \mathcal{B}_A$. So, if we set $S = S' \cap \mathcal{B}_A$, then $T_P(J)(A) = T_P(I)(A)$ for all $J \in \bigcap_{B \in S} \mathcal{G}(B)$ which is to say that T_P is locally finite for A and I. Since A and I were chosen arbitrarily, we obtain that T_P is locally finite. ∎

Exercise 21
Show that T_P is continuous in Q if P does not contain function symbols of arity greater than 0.

Exercise 22
Given a logic program P, a local variable *is a variable occurring in a body atom but not in the corresponding head. A program is* covered *if it does not contain local variables.*

(a) *Show that T_P is continuous in Q whenever P is covered.*

(b) *Does the converse of (a) also hold?*

Exercise 23
Let P be a logic program and let l be an injective level mapping for P. Show the following: If for each $A \in B_P$ there exists $n \in \mathbb{N}$ such that $l(B) \leq n$ for all $B \in \mathcal{B}_A$, then T_P is continuous in Q.

2.10 Proposition Let $I, M \in I_P$. If $T_P^n(I) \to M$ in Q, then M is a model of P. If furthermore T_P is locally finite, then M is a supported model of P (i.e. a fixed point of T_P).

Proof: Suppose $T_P^n(I) \to M$ in Q for some $I \in I_P$. We have to show that $T_P(M) \subseteq M$. Let $A \in T_P(M)$. By definition of T_P, there exists a ground instance $A \leftarrow A_1, \dots, A_{k_1}, \neg B_1, \dots, \neg B_{l_1}$ of a clause in P with $A_k \in M$ and $B_l \notin M$ for $k = 1, \dots, k_1$, $l = 1, \dots, l_1$. By Proposition 1.16, there is an $n_0 \in \mathbb{N}$, such that for all $n \geq n_0$, $A_k \in T_P^n(I)$ and $B_l \notin T_P^n(I)$ for all k, l. By definition of T_P and the above clause we have that $A \in T_P^m(I)$ for all $m \geq n_0 + 1$. Hence, $A \in T_P^n(I)$ eventually and therefore, by Proposition 1.16 again, $A \in M$, which proves the first statement.

The second statement follows by Proposition 1.18. ■

Exercise 24
Show the following: If I is a supported model of P, then there exists K such that $T_P^n(K) \to I$ in Q.

Exercise 25
Does the following hold? If M is a model of P then there exists K such that $T_P^n(K) \to I$ in Q.

2.11 Proposition Let P be a normal logic program and let $I_0 \in I_P$ be such that the sequence (I_n), with $I_n = T_P^n(I_0)$, converges in Q to some $M \in I_P$. If, for every $A \in M$, no clause whose head matches A contains a local variable, then M is a supported model.

Proof: We have to show that $M \subseteq T_P(M)$. So let $A \in M$. By convergence in Q and Proposition 1.16, there exists $n_0 \in \mathbb{N}$ such that $A \in T_P^n(I_0)$ for all $n \geq n_0$. By hypothesis, there are only finitely many ground instances of clauses in P with head A. Let C_0 be the (finite) set of all atoms occuring in positive body literals and D_0 the (finite) set of all atoms occuring in negative body literals of those clauses. Let $C_1 = C_0 \cap M$ and $D_1 = D_0 \setminus M$. Since $I_n \to M$ in Q, there is an $n_1 \in \mathbb{N}$ such that $C_1 \subseteq I_n$ and $D_1 \subseteq B_P \setminus I_n$ for all $n \geq n_1$. Since $A \in T_P(I_{\max\{n_0,n_1\}})$, there is a ground instance $A \leftarrow A_1, \dots, A_{k_1}, \neg B_1, \dots, \neg B_{l_1}$ of a clause in P with $A_k \in C_1 \subseteq M$ and $B_l \notin D_1 \subseteq B_P \setminus M$ for $k = 1, \dots, k_1$, $l = 1, \dots, l_1$. Hence $A \in T_P(M)$ as required. ■

2.3 Acyclic programs and recurrent networks

2.12 Definition A logic program P is *acyclic* if there exists a level mapping $l : B_P \to \mathbb{N}$ such that $l(A) > l(B)$ for all $B \in \mathcal{B}(A)$.

Exercise 26
Show that the programs

```
even(0).
even(s(X)) :- \+ even(s(X)).
```

and

```
member(X,[X|T]).
member(X,[H|T]) :- member(X,T).
```

are acyclic.

2.13 Definition Let P be a program, let l be a level mapping for P. The *Fitting metric* d_F on I_P is defined by $d_F(I,K) = 2^{-n}$, where $I, K \in I_P$ and n is smallest such that I and K differ on some atom of level n.

Exercise 27
Show that d_F is an ultrametric.

Exercise 28
Show that the Fitting metric induces Q if the level mapping is injective. Show that this does not hold in general if the level mapping is not injective.

Exercise 29
Show that if P is acyclic, then $T_P^n(I)$ converges in Q for all I.

For the following, let ι be a homeomorphism from (I_P, Q) onto Cantor space. Let P be a logic program such that T_P is continuous in Q, and denote $\iota(T_P)$ by F. Let f denote the input-output function of a corresponding network with error ε, in the sense of Theorem 2.4.

Exactly as for propositional programs, we want to endow the feedforward network resulting from Theorem 2.4 with recursive links, such that F can be iterated.

Assume furthermore for the following that F is *Lipschitz-continuous*, that is, there exists $\lambda \geq 0$ such that for all $x, y \in \mathcal{C}$ we have $d(F(x), F(y)) \leq \lambda d(x,y)$.

Exercise 30
Show that Lipschitz-continuity implies continuity in any metric space.

For $x, y \in \mathcal{C}$ we obtain

$$d(f(x), F(y)) \leq d(f(x), F(x)) + d(F(x), F(y)) \leq \varepsilon + \lambda d(x,y). \tag{1}$$

Now let $x \in \mathcal{C}$ be arbitrarily chosen. By Equation (1) we obtain

$$d\left(f^2(x), F^2(x)\right) \leq \varepsilon + \lambda d(f(x), F(x)) \leq \varepsilon + \lambda\varepsilon. \tag{2}$$

Inductively, we can prove that for all $n \in \mathbb{N}$ we have

$$d\left(f^n(x), F^n(x)\right) \leq \varepsilon + \lambda\varepsilon + \cdots + \lambda^{n-1}\varepsilon = \varepsilon\left(\sum_{i=0}^{n-1} \lambda^i\right) = \varepsilon\frac{1-\lambda^n}{1-\lambda}. \tag{3}$$

Thus, we obtain the following bound on the error produced by the recurrent network after n iterations, assuming that d is the natural metric on $\mathbb{R}$.

2.14 Theorem With the notation and hypotheses above, for any $I \in I_P$ and any $n \in \mathbb{N}$ we have

$$|f^n(\iota(I)) - \iota(T_P^n(I))| \leq \varepsilon \frac{1-\lambda^n}{1-\lambda}.$$

Proof: Note that $\iota(T_P^n(I)) = F^n(\iota(I))$, and the assertion follows from Equation (3) since d is the natural metric on $\mathbb{R}$. ■

2.15 Corollary If F is a contraction on $\mathcal{C}$, so that $\lambda < 1$, then $\left(F^k(\iota(I))\right)$ converges for every I to the unique fixed point x of F and there exists $m \in \mathbb{N}$ such that for all $n \geq m$ we have

$$|f^n(\iota(I)) - x| \leq \varepsilon \frac{1}{1-\lambda}.$$

Proof: The convergence follows from the Banach contraction mapping theorem. The inequality follows immediately from Theorem 2.14 using the expression for limits of geometric series as in the proof of Proposition 1.10. ■

2.16 Remark It was shown in [HKS99], that we can assure the hypotheses of Corollary 2.15 as follows. We embed I_P into $\mathbb{R}$ by using the quaternary number system with digits 0 and 1, i.e. we use the representation of Cantor set as the set of all real numbers which can be written in the form $\sum_{i=1}^{\infty} a_i 4^{-i}$, where $a_i \in \{0, 1\}$ for all i. It was shown in [HKS99], that for every program which is acyclic with respect to the bijective level mapping used for the embedding ι, we have that $\iota(T_P)$ is a contraction on $\mathcal{C}$ with respect to the natural metric from $\mathbb{R}$. Thus, in this case the hypotheses of Corollary 2.15 are satisfied.

2.17 Remark In [BH04] an entirely different approach was taken to representing first-order logic programs as artificial neural networks. The resulting architecture is entirely different, very specific, but uses standard Gaussian activation functions. Continuity of T_P with respect to the Cantor topology, and Lipschitz-continuity of $\iota(T_P)$ also play a major role there.

2.18 Research Problem Can we cast the results from Remark 2.17 into an integrated neural-symbolic learning system?

3 Further reading

[Wil70] is an excellent book on set-theoretic topology. The Cantor topology was introduced to logic programming in [Bat89, BS89b, BS89a] and further studied in [Sed95, Sed97]. It appears basically every time when iterative behaviour of T_P for normal P is being studied, one may want to consult [Hit01, HS03]. There is a result due to Hornik, Stinchcombe and White [HSW89], which generalizes Funahashi's (Theorem 2.3) in that it provides an approximation result for *measurable* functions. The implications of this remain to be worked out, and this is briefly discussed in [HHS0x].

References

[Bat89] Aida Batarekh. *Topological Aspects of Logic Programming.* PhD thesis, Syracuse University, June 1989.

[BH04] Sebastian Bader and Pascal Hitzler. Logic programs, iterated function systems, and recurrent radial basis function networks. *Journal of Applied Logic*, 2004. To appear.

[BS89a] Aida Batarekh and V.S. Subrahmanian. The query topology in logic programming. In *Proceedings of the 1989 Symposium on Theoretical Aspects of Computer Science*, volume 349 of *Lecture Notes in Computer Science*, pages 375–387. Springer, Berlin, 1989.

[BS89b] Aida Batarekh and V.S. Subrahmanian. Topological model set deformations in logic programming. *Fundamenta Informaticae*, 12:357–400, 1989.

[Fun89] Ken-Ichi Funahashi. On the approximate realization of continuous mappings by neural networks. *Neural Networks*, 2:183–192, 1989.

[HHS0x] Pascal Hitzler, Steffen Hölldobler, and Anthony K. Seda. Logic programs and connectionist networks. *Journal of Applied Logic*, 200x. To appear.

[Hit01] Pascal Hitzler. *Generalized Metrics and Topology in Logic Programming Semantics.* PhD thesis, Department of Mathematics, National University of Ireland, University College Cork, 2001.

[HKS99] Steffen Hölldobler, Yvonne Kalinke, and Hans-Peter Störr. Approximating the semantics of logic programs by recurrent neural networks. *Applied Intelligence*, 11:45–58, 1999.

[HS00] Pascal Hitzler and Anthony K. Seda. A note on relationships between logic programs and neural networks. In Paul Gibson and David Sinclair, editors, *Proceedings of the Fourth Irish Workshop on Formal Methods, IWFM'00*, Electronic Workshops in Comupting (eWiC). British Computer Society, 2000.

[HS03] Pascal Hitzler and Anthony K. Seda. Generalized metrics and uniquely determined logic programs. *Theoretical Computer Science*, 305(1–3):187–219, 2003.

[HSW89] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2:359–366, 1989.

[Sed95] Anthony K. Seda. Topology and the semantics of logic programs. *Fundamenta Informaticae*, 24(4):359–386, 1995.

[Sed97] Anthony K. Seda. Quasi-metrics and the semantics of logic programs. *Fundamenta Informaticae*, 29(1):97–117, 1997.

[Wil70] Stephen Willard. *General Topology.* Addison-Wesley, Reading, MA, 1970.