Neural Networks Between Integer and Rational Weights

Jiří Šíma

Institute of Computer Science
The Czech Academy of Sciences

The Neural Network Model

- Architecture: s computational units (neurons), indexed as $V = \{1, \ldots, s\}$, connected into a directed graph (V, A) where $A \subseteq V \times V$

- each edge $(i, j) \in A$ from unit i to j is labeled with a real weight $w_{ji} \in \mathbb{R}$ ($w_{ji} = 0$ iff $(i, j) \notin A$)

- each neuron $j \in V$ is associated with a real bias $w_{j0} \in \mathbb{R}$ (i.e. a weight of $(0, j) \in A$ from an additional neuron $0 \in V$)

- Computational Dynamics: the evolution of network state (output)

$$\mathbf{y}^{(t)} = (y_1^{(t)}, \ldots, y_s^{(t)}) \in [0, 1]^s$$

  at discrete time instant $t = 0, 1, 2, \ldots$

Discrete-Time Computational Dynamics

1. initial state $\mathbf{y}^{(0)} \in [0,1]^s$

2. at discrete time instant $t \geq 0$, an excitation is computed as

$$\xi_j^{(t)} = w_{j0} + \sum_{i=1}^{s} w_{ji} y_i^{(t)} = \sum_{i=0}^{s} w_{ji} y_i^{(t)} \qquad \text{for } j = 1, \ldots, s$$

where unit $0 \in V$ has constant output $y_0^{(t)} \equiv 1$ for every $t \geq 0$

Discrete-Time Computational Dynamics (continued)

3. at the next time instant $t+1$, only the neurons $j \in \alpha_{t+1}$ from a selected subset $\alpha_{t+1} \subseteq V$ update their states:

$$y_j^{(t+1)} = \begin{cases} \sigma(\xi_j) & \text{for } j \in \alpha_{t+1} \\ y_j^{(t)} & \text{for } j \in V \setminus \alpha_{t+1} \end{cases}$$

where $\sigma : \mathbb{R} \longrightarrow [0,1]$ is an activation function, e.g.

$$\sigma(\xi) = \begin{cases} 1 & \text{for } \xi \geq 1 \\ \xi & \text{for } 0 < \xi < 1 \\ 0 & \text{for } \xi \leq 0 \end{cases}$$ the saturated-linear function

Neural Networks as Language Acceptors

- language (problem) $L \subseteq \Sigma^*$ over a finite alphabet Σ

- input string $x_1 \ldots x_n \in \Sigma^n$ of arbitrary length $n \geq 0$ is sequentially presented, symbol after symbol, via input neurons $i \in X = \text{enum}(\Sigma) \subseteq V$:

$$y_i^{(d(\tau-1))} = \begin{cases} 1 & \text{for } i = \text{enum}(x_\tau) \\ 0 & \text{for } i \neq \text{enum}(x_\tau) \end{cases} \quad \text{at macroscopic time } \tau = 1, \ldots, n$$

where integer $d \geq 1$ is the time overhead for processing a single input symbol

- output neuron out $\in V$ signals whether input $x_1 \ldots x_n \overset{?}{\in} L$:

$$y_{\text{out}}^{(T(n))} = \begin{cases} 1 & \text{if } x_1 \ldots x_n \in L \\ 0 & \text{if } x_1 \ldots x_n \notin L \end{cases}$$

where $T(n)$ is the computational time in terms of input length n

The Computational Power of Neural Networks

depends on the information contents of weight parameters:

1. integer weights: finite automaton `(Minsky, 1967)`

2. rational weights: Turing machine `(Siegelmann, Sontag, 1995)`
   polynomial time $\equiv$ complexity class P

3. arbitrary real weights: "super-Turing" computation `(Siegelmann, Sontag, 1994)`
   polynomial time $\equiv$ nonuniform complexity class P/poly
   exponential time $\equiv$ any I/O mapping

The Computational Power of Neural Networks

depends on the information contents of weight parameters:

1. integer weights: finite automaton (Minsky, 1967)

2. rational weights: Turing machine (Siegelmann, Sontag, 1995)
   polynomial time $\equiv$ complexity class P

   polynomial time & increasing Kolmogorov complexity of real weights $\equiv$
   a proper hierarchy of nonuniform complexity classes between P and P/poly
   (Balcázar, Gavaldà, Siegelmann, 1997)

3. arbitrary real weights: "super-Turing" computation (Siegelmann, Sontag, 1994)
   polynomial time $\equiv$ nonuniform complexity class P/poly
   exponential time $\equiv$ any I/O mapping

The Computational Power of Neural Networks

depends on the information contents of weight parameters:

1. integer weights: finite automaton (Minsky, 1967)

   a gap between integer a rational weights w.r.t. the Chomsky hierarchy

   regular (Type-3) $\times$ recursively enumerable (Type-0) languages

2. rational weights: Turing machine (Siegelmann, Sontag, 1995)
   polynomial time $\equiv$ complexity class P

   polynomial time & increasing Kolmogorov complexity of real weights $\equiv$
   a proper hierarchy of nonuniform complexity classes between P and P/poly
   (Balcázar, Gavaldà, Siegelmann, 1997)

3. arbitrary real weights: "super-Turing" computation (Siegelmann, Sontag, 1994)
   polynomial time $\equiv$ nonuniform complexity class P/poly
   exponential time $\equiv$ any I/O mapping

Neural Networks Between Integer and Rational Weights

TWO analog neurons with rational weights $(+$ a few integer-weight neurons) can implement a 2-stack pushdown automaton $\equiv$ Turing machine

$\longrightarrow$ **What is the computational power of ONE extra analog neuron ?**

A Neural Network with an Extra Analog Neuron (NN1A):

all the weights to neurons are integers except for ONE neuron s with rational weights:

$$w_{ji} \in \begin{cases} \mathbb{Z} & j = 1, \ldots, s-1 \\ \mathbb{Q} & j = s \end{cases} \quad , \quad i \in \{0, \ldots, s\}$$

or equivalently: rational weights $+$ all the neurons but ONE analog unit employ the Heaviside activation function:

$$\sigma_j(\xi) = \begin{cases} 1 & \text{for } \xi \geq 0 \\ 0 & \text{for } \xi < 0 \end{cases} \qquad j = 1, \ldots, s-1$$

The Representation Theorem for NN1A

(Our Main Technical Result)

a language $L \subset \Sigma^*$ that is accepted by a NN1A, can be written as

$$L = h\left(\left(\left(\bigcup_{r=1}^{p-1}\left(\overline{L_{<c_r}} \cap L_{<c_{r+1}}\right) \cdot A_r\right)^{Pref} \cap R_0\right)^* \cap R\right)$$

(options: $\overline{L_{>0}}\,,\ L_{>c_r} \cap L_{<c_{r+1}}\,,\ L_{>c_r} \cap \overline{L_{>c_{r+1}}}\,,\ \overline{L_{<c_r}} \cap \overline{L_{<c_{r+1}}}\,,\ \overline{L_{<1}}$)

where

- $A_1, \ldots, A_p$ is a partition of a finite alphabet A
- S^{Pref} denotes the largest prefix-closed subset of $S \cup A \cup \{\varepsilon\}$
- $R\,, R_0 \subseteq A^*$ are regular languages
- $h : A^* \longrightarrow \Sigma^*$ is a letter-to-letter morphism
- $L_{<c_r}\,,\ L_{>c_r} \subseteq A^*$ are so-called **cut languages** for rational thresholds

$$0 = c_1 \leq c_2 \leq \cdots \leq c_p = 1$$

Cut Languages

a cut language $L_{<c}$ contains all the finite β-expansions $a_1 \ldots a_n \in A^*$ of numbers that are less than a threshold $c \in \mathbb{R}$ (similarly for $L_{>c}$):

$$L_{<c} = \left\{ a_1 \ldots a_n \in A^* \;\middle|\; (0 \,.\, a_1 \ldots a_n)_\beta = \sum_{k=1}^{n} a_k \beta^{-k} < c \right\}$$

β-expansions: base-β representations of numbers using digits from A where

- $\beta \in \mathbb{R}$ is a base (radix) such that $|\beta| > 1$
- $\emptyset \neq A \subset \mathbb{R}$ is a finite set of digits

a generalization of integer-base positional numeral systems, e.g.

- decimal expansions: $\beta = 10$ and $A = \{0, 1, 2, \ldots, 9\}$
  e.g. $\frac{3}{4} = (0\,.\,75)_{10} = 7 \cdot 10^{-1} + 5 \cdot 10^{-2}$
- binary expansions: $\beta = 2$ and $A = \{0, 1\}$
  e.g. $\frac{3}{4} = (0\,.\,11)_2 = 1 \cdot 2^{-1} + 1 \cdot 2^{-2}$

Infinite β-Expansions (Rényi, 1957; Parry, 1960)

an infinite word $a_1a_2a_3\cdots \in A^\omega$ is a β-expansion of number

$$(0\,.\,a_1a_2a_3\cdots)_\beta = \sum_{k=1}^{\infty} a_k\beta^{-k}$$

Uniqueness:

1. for integer base $\beta > 0$ and $A = \{0, 1, \ldots, \beta - 1\}$, any number from $[0, 1]$ has a unique β-expansion except for those with finite β-expansions,

   e.g. $\frac{3}{4} = (0\,.\,75)_{10} = (0\,.\,75000\ldots)_{10} = (0\,.\,74999\ldots)_{10}$

2. for non-integer base β, almost every number has a continuum of distinct β-expansions (Sidorov, 2003)

   Example: $1 < \beta < 2$, $A = \{0, 1\}$, $D_\beta = \left(0\,,\,\frac{1}{\beta-1}\right)$

   - $1 < \beta < \varphi$ where $\varphi = (1+\sqrt{5})/2 \approx 1.618034$ is the golden ratio: every $x \in D_\beta$ has a continuum of distinct β-expansions (Erdös et al., 1990)
   - $\varphi \leq \beta < q$ where $q \approx 1.787232$ is the Komornik-Loreti constant: countably many $x \in D_\beta$ have unique β-expansions (Glendinning,Sidorov,2001)
   - $q \leq \beta < 2$: a continuum of $x \in D_\beta$ with unique β-expansions

Eventually Periodic β-Expansions

$$a_1 a_2 \dots a_{k_1} \, (a_{k_1+1} a_{k_1+2} \dots a_{k_2})^\omega$$

- $a_1 a_2 \dots a_{k_1} \in A^{k_1}$ is a preperiodic part of length k_1 (purely periodic β-expansions satisfy $k_1 = 0$)

- $a_{k_1+1} a_{k_1+2} \dots a_{k_2} \in A^m$ is a repetend of length $m = k_2 - k_1 > 0$ whose minimum is the period of β-expansion

- $(0 \,.\, a_1 a_2 \dots a_{k_1} \overline{a_{k_1+1} a_{k_1+2} \dots a_{k_2}})_\beta = (0 \,.\, a_1 a_2 \dots a_{k_1})_\beta + \beta^{-k_1} \varrho$

  where $\varrho = (0 \,.\, \overline{a_{k_1+1} a_{k_1+2} \dots a_{k_2}})_\beta = \dfrac{\sum_{k=1}^{m} a_{k_1+k} \beta^{-k}}{1 - \beta^{-m}}$ is a periodic point

Example: $\beta = \frac{3}{2}$, $A = \{0, 1\}$, $1\,(10)^\omega = 1\,10\,10\,10\,10\,10\,\dots$

$$\frac{22}{15} = (0 \,.\, 1\, \overline{10})_{\frac{3}{2}} = (0 \,.\, 1)_{\frac{3}{2}} + \left(\frac{3}{2}\right)^{-1} \cdot \varrho \quad \text{where} \quad \varrho = \frac{\left(\frac{3}{2}\right)^{-1}}{1 - \left(\frac{3}{2}\right)^{-2}} = \frac{6}{5}$$

Eventually Quasi-Periodic β-Expansions

β-expansion $a_1 \ldots a_{k_1} \, a_{k_1+1} \ldots a_{k_2} \, a_{k_2+1} \ldots a_{k_3} \, a_{k_3+1} \ldots a_{k_4} \ldots \in A^\omega$

is eventually quasi-periodic if there is $0 \le k_1 < k_2 < \cdots$ such that

$$\varrho = (0\,.\,\overline{a_{k_1+1} \ldots a_{k_2}})_\beta = (0\,.\,\overline{a_{k_2+1} \ldots a_{k_3}})_\beta = (0\,.\,\overline{a_{k_3+1} \ldots a_{k_4}})_\beta = \cdots$$

- $a_1 a_2 \ldots a_{k_1} \in A^{k_1}$ is a preperiodic part of length k_1 (purely quasi-periodic β-expansions satisfy $k_1 = 0$)
- $a_{k_i+1} \ldots a_{k_{i+1}} \in A^{m_i}$ is a quasi-repetend of length $m_i = k_{i+1} - k_i > 0$
- $(0\,.\,a_1 a_2 a_2 \ldots)_\beta = (0\,.\,a_1 a_2 \ldots a_{k_1})_\beta + \beta^{-k_1} \varrho$ where for every $i \ge 1$,

$$(0\,.\,\overline{a_{k_i+1} \ldots a_{k_{i+1}}})_\beta = \frac{\sum_{k=1}^{m_i} a_{k_i+k} \beta^{-k}}{1 - \beta^{-m_i}} = \varrho \quad \text{is a periodic point}$$

$\longrightarrow$ quasi-repetends can be mutually replaced with each other arbitrarily

- a generalization of eventually periodic β-expansions:

$$a_{k_1+1} \ldots a_{k_2} = a_{k_2+1} \ldots a_{k_3} = a_{k_3+1} \ldots a_{k_4} = \cdots$$

An Example of Quasi-Periodic β-Expansion

base $\boldsymbol{\beta = \frac{5}{2}}$, digits $\boldsymbol{A = \left\{0\,,\,\frac{1}{2}\,,\,\frac{7}{4}\right\}}$, periodic point $\boldsymbol{\varrho = \frac{3}{4}}$

$$\left(0\,.\,\overline{\tfrac{7}{4}\;0}\right)_{\frac{5}{2}} = \left(0\,.\,\overline{\tfrac{7}{4}\;\tfrac{1}{2}\;0}\right)_{\frac{5}{2}} = \left(0\,.\,\overline{\tfrac{7}{4}\;\tfrac{1}{2}\;\tfrac{1}{2}\;0}\right)_{\frac{5}{2}} = \left(0\,.\,\overline{\tfrac{7}{4}\;\tfrac{1}{2}\;\tfrac{1}{2}\;\tfrac{1}{2}\;0}\right)_{\frac{5}{2}}$$

$$= \cdots = \left(0\,.\,\overline{\tfrac{7}{4}\;\underbrace{\tfrac{1}{2}\;\cdots\;\tfrac{1}{2}}_{n\text{ times}}\;0}\right)_{\frac{5}{2}} = \cdots = \frac{\frac{7}{4}\left(\frac{5}{2}\right)^{-1} + \sum_{i=2}^{n+1} \frac{1}{2}\cdot\left(\frac{5}{2}\right)^{-i}}{1-\left(\frac{5}{2}\right)^{-n-2}} = \frac{3}{4}$$

$\longrightarrow$ $\boldsymbol{\varrho = \frac{3}{4}}$ has uncountably many distinct quasi-periodic β-expansions:

$$\frac{3}{4} = \left(0\,.\,\tfrac{7}{4}\;\underbrace{\tfrac{1}{2}\;\cdots\;\tfrac{1}{2}}_{n_1\text{ times}}\;0\;\tfrac{7}{4}\;\underbrace{\tfrac{1}{2}\;\cdots\;\tfrac{1}{2}}_{n_2\text{ times}}\;0\;\tfrac{7}{4}\;\underbrace{\tfrac{1}{2}\;\cdots\;\tfrac{1}{2}}_{n_3\text{ times}}\;0\;\tfrac{7}{4}\;\underbrace{\tfrac{1}{2}\;\cdots\;\tfrac{1}{2}}_{n_4\text{ times}}\;0\;\cdots\right)_{\frac{5}{2}}$$

where $\boldsymbol{n_1, n_2, n_3, \ldots}$ is any infinite sequence of nonnegative integers

Quasi-Periodic Numbers

$c \in \mathbb{R}$ is β-quasi-periodic within A if every infinite β-expansion of c is eventually quasi-periodic

Examples:

- c from the complement of the Cantor set **is** 3-quasi-periodic within $\{0, 2\}$:
  c has no β-expansion at all

- $c = \frac{3}{4}$ **is** $\frac{5}{2}$-quasi-periodic within $A = \{0\,,\,\frac{1}{2}\,,\,\frac{7}{4}\}$:
  all the $\frac{5}{2}$-expansions of $\frac{3}{4}$ using digits from A, are eventually quasi-periodic

- $c = \frac{40}{57} = (0\,.\,0\,\overline{011})_{\frac{3}{2}}$ **is not** $\frac{3}{2}$-quasi-periodic within $A = \{0, 1\}$:
  greedy (i.e. lexicographically maximal) $\frac{3}{2}$-expansion $100000001\ldots$ of $\frac{40}{57}$ is not eventually periodic

Cut Languages Within the Chomsky Hierarchy

(Šíma, Savický, LATA 2017)

Theorem 1 *A cut language $L_{<c}$ is regular* **iff** *c is β-quasi-periodic within A.*

Example: any regular language $L \subset A^*$ where $\{\alpha_1, \alpha_2\} \subseteq A$ such that $L \cap \{\alpha_1, \alpha_2\}^2 = \{\alpha_1\alpha_2, \alpha_2\alpha_1\}$, is not a cut language

Theorem 2 *Let $\beta \in \mathbb{Q}$ and $A \subset \mathbb{Q}$. Every cut language $L_{<c}$ with threshold $c \in \mathbb{Q}$ is context-sensitive.*

Theorem 3 *If c is not β-quasi-periodic within A, then the cut language $L_{<c}$ is not context-free.*

Corollary 1 *Any cut language $L_{<c}$ is either regular or non-context-free (depending on whether c is a β-quasi-periodic number within A).*

The Computational Power of NN1A

the results on cut languages + **representation theorem** for NN1A:

$$L = h\left(\left(\left(\bigcup_{r=1}^{p-1}\left(\overline{L_{<c_r}} \cap L_{<c_{r+1}}\right) \cdot A_r\right)^{Pref} \cap R_0\right)^* \cap R\right)$$

where $\beta = \frac{1}{w_{ss}}$, $A = \left\{ \sum_{i=0}^{s-1} w_{si} y_i \,\middle|\, y_1, \ldots, y_{s-1} \in \{0,1\} \right\} \cup \{0,1\}$,

$$C = \{c_1, \ldots, c_p\} = \left\{ -\sum_{i=0}^{s-1} \frac{w_{ji}}{w_{js}} y_i \,\middle|\, j \in V \setminus (X \cup \{s\}) \text{ s.t. } w_{js} \neq 0\,, \; y_1, \ldots, y_{s-1} \in \{0,1\} \right\} \cup \{0,1\}$$

Theorem 4 *Let N be a NN1A and assume $0 < |w_{ss}| < 1$. Define $\beta \in \mathbb{Q}$, $A \subset \mathbb{Q}$, and $C \subset \mathbb{Q}$ using the weights of N. If every $c \in C$ is β-quasi-periodic within A, then N accepts regular language.*

Theorem 5 *There is a language accepted by a NN1A, which is not context-free.*

Theorem 6 *Any language accepted by a NN1A is context-sensitive.*

Conclusions

- we have characterized the class of languages accepted by NN1As—integer-weight neural networks with an extra rational-weight neuron, using cut languages
- we have shown an interesting link to active research on β-expansions in non-integer bases
- we have refined the analysis of the computational power of neural networks between integer and rational weights within the Chomsky hierarchy:

  integer-weight NNs $\equiv$ regular languages (Type 3)

  a sufficient condition when NN1A accepts regular language (Type 3)

  a language accepted by NN1A that is not context-free (Type 2)

  NN1As $\subset$ context-sensitive languages (Type 1)

  rational-weight NNs $\equiv$ recursively enumerable languages (Type 0)

- **Open problems:**
  - a necessary condition when NN1A accepts a regular language
  - the analysis for $w_{ss} \in \mathbb{R}$ or $|w_{ss}| > 1$
  - a proper hierarchy of NNs e.g. with increasing quasi-period of weights