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)

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

  at discrete time instant $t = 0, 1, 2, \ldots$
Discrete-Time Computational Dynamics

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

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

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

where unit $0 \in \mathcal{V}$ 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 $\sigma : \mathbb{R} \rightarrow [0, 1]$ 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 $\Sigma$

- 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** $\text{out} \in V$ signals whether input $x_1 \ldots x_n \in L$:

$$y_{\text{out}}^{(T(n))} = \begin{cases} 1 & \text{if } x_1 \ldots x_n \in L \\ 0 & \text{if } x_1 \ldots x_n \not\in L \end{cases}$$

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) × recursively enumerable (Type-0) languages

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

   polynomial time & increasing Kolmogorov complexity of real weights ≡
   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 ≡ nonuniform complexity class P/poly
   exponential time ≡ 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

\[ \rightarrow \text{What is the computational power of ONE extra analog neuron?} \]

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

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

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

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

\[ \sigma_j(\xi) = \begin{cases} 1 & \text{for } \xi \geq 0 \\ 0 & \text{for } \xi < 0 \end{cases}, \quad j = 1, \ldots, s - 1 \]
The Representation Theorem for NN1A  
(Our Main Technical Result)

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

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

(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
- \( h : A^* \rightarrow \Sigma^* \) 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 $\beta$-expansions $a_1 \ldots a_n \in A^*$ of numbers that are less than a threshold $c \in \mathbb{R}$ (similarly for $L_{>c}$):

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

**\(\beta\)-expansions:** Base-$\beta$ 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 $\beta$-Expansions (Rényi, 1957; Parry, 1960)

an infinite word $a_1a_2a_3\cdots \in A^\omega$ is a $\beta$-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 $\beta$-expansion except for those with finite $\beta$-expansions,
  
  e.g. $\frac{3}{4} = (0.75)_{10} = (0.75000\ldots)_{10} = (0.74999\ldots)_{10}$

2. for non-integer base $\beta$, almost every number has a continuum of distinct $\beta$-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 $\beta$-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 $\beta$-expansions (Glendinning, Sidorov, 2001)
- $q \leq \beta < 2$: a continuum of $x \in D_\beta$ with unique $\beta$-expansions
Eventually Periodic $\beta$-Expansions

$$a_1a_2 \ldots a_{k_1} (a_{k_1+1}a_{k_1+2} \ldots a_{k_2})^\omega$$

- $a_1a_2 \ldots a_{k_1} \in A^{k_1}$ is a preperiodic part of length $k_1$
  (purely periodic $\beta$-expansions satisfy $k_1 = 0$)

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

- $(0 \cdot a_1a_2 \ldots a_{k_1} a_{k_1+1}a_{k_1+2} \ldots a_{k_2}) \beta = (0 \cdot a_1a_2 \ldots a_{k_1}) \beta + \beta^{-k_1} \varrho$

  where $\varrho = (0 \cdot a_{k_1+1}a_{k_1+2} \ldots a_{k_2}) \beta = \frac{\sum_{k=1}^m a_{k_1+k} \beta^{-k}}{1 - \beta^{-m}}$ is a periodic point

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

$$\frac{22}{15} = (0.1\overline{10})_{\frac{3}{2}} = (0.1)_\frac{3}{2} + \left(\frac{3}{2}\right)^{-1} \cdot \varrho$$

where $\varrho = \frac{(\frac{3}{2})^{-1}}{1 - (\frac{3}{2})^{-2}} = \frac{6}{5}$
Eventually Quasi-Periodic $\beta$-Expansions

$\beta$-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 $0 \leq k_1 < k_2 < \cdots$ 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 $\beta$-expansions satisfy $k_1 = 0$)

- $a_{k_i+1} \ldots a_{k_{i+1}} \in A^{m_i}$ is a quasi-repetend of length $m_i = k_{i+1} - k_i > 0$

- $(0.\overline{a_1 a_2 a_2 \ldots})_{\beta} = (0.\overline{a_1 a_2 \ldots a_{k_1}})_{\beta} + \beta^{-k_1} \varrho$ where for every $i \geq 1$,

  $$(0.\overline{a_{k_i+1} \ldots a_{k_{i+1}}})_{\beta} = \frac{\sum_{k=1}^{m_i} a_{k_i+k} \beta^{-k}}{1 - \beta^{-m_i}} = \varrho$$

  is a periodic point

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

- a generalization of eventually periodic $\beta$-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 $\beta$-Expansion

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

$\left(0. \frac{7}{4} 0\right)^{\frac{5}{2}} = \left(0. \frac{7}{4} 2 0\right)^{\frac{5}{2}} = \left(0. \frac{7}{4} 2 2 0\right)^{\frac{5}{2}} = \left(0. \frac{7}{4} 2 2 2 0\right)^{\frac{5}{2}}$

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

$\longrightarrow \rho = \frac{3}{4}$ has uncountably many distinct quasi-periodic $\beta$-expansions:

$\frac{3}{4} = \left(0. \frac{7}{4} 2 \cdots \frac{1}{2} 0 \frac{7}{4} 2 \cdots \frac{1}{2} 0 \frac{7}{4} 2 \cdots \frac{1}{2} 0 \frac{7}{4} 2 \cdots \frac{1}{2} 0 \cdots \right)^{\frac{5}{2}}$

where $n_1, n_2, n_3, \ldots$ is any infinite sequence of nonnegative integers
Quasi-Periodic Numbers

c ∈ \mathbb{R} is $\beta$-quasi-periodic within \( A \) if every infinite $\beta$-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 $\beta$-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... 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 $\beta$-quasi-periodic within $A$.

**Example:** any regular language $L \subset A^*$ where $\{\alpha_1, \alpha_2\} \subset A$ 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 $\beta$-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 $\beta$-quasi-periodic number within $A$).
The Computational Power of NN1A

the results on cut languages + representation theorem for NN1A:

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

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

  $\quad$ 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