Parsimonious Neural Network Abduction

Preprint · June 2025

DOI: 10.13140/RG.2.2.15695.80804

Gary Nan Tie

Mu Risk LLC

Parsimonious Neural Network Abduction

Gary Nan Tie, Jun 22, 2025

Abstract

For hypotheses whose effect is manifested by observations, abduction seeks to explain a given observation by finding a hypothesis whose effect is that observation.

Abductive reasoning has a fibration semantics [1] that can be implemented by neural networks [2]; a step towards artificial general intelligence.

In this note we introduce a parsimonious choice of explanation depending on one's utility function.

An arrow $p : K \to V$ in a category is said to be a <u>fibration</u> with respect to a class of arrows $\mathcal{M}$ when for any $m \in \mathcal{M}$ and 2-cell $m \overset{(f,g)}{\Rightarrow} p$, that is solid commutative square

$$
\begin{tikzcd}
B \arrow[r, "f"] \arrow[d, "m"'] & K \arrow[d, "p"] \\
Q \arrow[r, "g"'] \arrow[ur, dotted, "\sigma"] & V
\end{tikzcd}
$$

there exists dotted lifting σ making both triangles commute.

Now let the category arrows be neural networks on finite data sets. Learn $I : V \to K$ from I/O $\{(g(q), \sigma(q))\}_{q \in Q}$ subject to $I \circ g = \sigma$, then $(p \circ I) \circ g \approx g$, that is $p \circ I|_{g(Q)} \approx 1_{g(Q)}$, which is to say, on $g(Q) \subseteq V$, effect p is explained by abduction I, when p is a fibration, (and assume lifting σ is learnable).

Let analogies $\mathcal{A}_{p,g}$ be a finite subset of

$$\{ m \overset{(f,g)}{\Longrightarrow} p \mid m \in \mathcal{M},\ \text{span}(m,f),\ p \text{ a fibration wrt } \mathcal{M} \}$$

with corresponding abductions

$$\mathcal{T} = \{ T : V \to K \mid T \circ g = \sigma,\ \sigma \text{ a lifting from } \mathcal{A} \}$$

Let $L : V \times V \to \mathbb{R}$ be your loss function.

For $T \in \mathcal{T}$, let average discrepancy

$$\mu_T \triangleq \frac{1}{|g(Q)|} \sum_{v \in g(Q)} L(p(T(v)), v)$$

Define $T \precsim T'$ iff $\mu_T \leq \mu_{T'}$

With respect to analogies $\mathcal{A}$

$$\tilde{T} \triangleq \min_{\precsim} \{ T \in \mathcal{T} \}$$

is our parsiomonius explanation of effect p.

Summary:

For neural networks
$$\begin{tikzcd} & K \arrow[d, "p \text{ effect}"] \\ Q \arrow[r, "g"', "\text{attention}"'] & V \end{tikzcd}$$

when p is a fibration with respect to $\mathcal{M}$,

learned liftings σ for analogies (2-cells),

induce abductions I. The abduction with the

least average discrepancy is chosen to our parsimonius

explanation of effect p, according to our utility.

References

[1] ‘Fibrations explain all you need!’
- Fibrations, Abduction and Attention
Gary Nan Tie, Mar 4, 2025
DOI: 10.13140/RG.2.2.35984.52488

[2] ‘Neural Network Abduction’
Gary Nan Tie, Jun 13, 2025
DOI: 10.13140/RG.2.2.18506.07360/1