Between Min Cost Search[1] and Least Action[2]

Tim Fernando

`Tim.Fernando@tcd.ie`

EuroProofNet-WG5, Edinburgh, April 2025

Trinity College Dublin
The University of Dublin

[1] Breadth-first search generalized to arc costs adding up along a path (Dijkstra's algorithm)

[2] $\mathcal{S}(\mathbf{q}) = \int_0^{\hat{t}} \mathcal{L}(\mathbf{q}(t), \dot{\mathbf{q}}(t))\, dt$ with Lagrangian $\mathcal{L}(q, \dot{q})$ (Hamilton's principle, $\delta\mathcal{S} = 0$)

Abstract

The function from a string $a_1 a_2 \cdots a_n$ of n words/tokens a_i to probability

$$P(a_1 a_2 \cdots a_n) = \prod_{i=0}^{n-1} P(a_{i+1} | q_i) \quad \text{where } q_i = a_1 a_2 \cdots a_i \tag{1}$$

is at its maximum when Shannon **surprisal**

$$-\log P(a_1 a_2 \cdots a_n) = \sum_{i=0}^{n-1} c_i \quad \text{where } c_i = -\log P(a_{i+1} | q_i) \tag{2}$$

is at its minimum, with arc costs c_i adding up along a path

$$q_0 \overset{a_1}{\to} q_1 \overset{a_2}{\to} q_2 \cdots \overset{a_n}{\to} q_n. \tag{3}$$

The assumption $q_i = a_1 a_2 \cdots a_i$ is relaxed below, and strings s other than $a_1 a_2 \cdots a_n$ are associated with paths (3) that s might record only sporadically and imperfectly. The strings s are

(1) compared to state sequences $\mathbf{q}(t_1)\mathbf{q}(t_2) \cdots \mathbf{q}(t_n)$ used to approximate the action integral $\int_0^{\hat{t}} \mathcal{L}(\mathbf{q}(t), \dot{\mathbf{q}}(t))\, dt$ in discrete mechanics (Marsden & West 2001),

(2) shaped by a signature Σ capturing Kleene's finite-state representation of McCulloch Pitts nerve nets,

(3) collected in a set $\mathbf{Mod}(\Sigma)$ by a contravariant functor **Mod** from a partial order on signatures Σ, paired with a functor *Sen* mapping Σ covariantly to a set $\mathit{Sen}(\Sigma)$ of Σ-sentences φ that s may (or may not) satisfy under a logical system called an *institution* (Goguen & Burstall 1992), and

(4) compressed by a function κ reducing time$_\Sigma$ to change$_\Sigma$ so that refinements of Σ may lengthen strings, leading to further pattern deformations (Mumford 1994).

The claim is that **a nascent notion of state in (1) can be refined to understand inference (not to mention pre-training or fine-tuning) in large language models as a min-cost search over paths deformed and observed partially as strings** s, on which finite-state constraints (hard and soft) are imposed below.

Step 1: a_i as a fragment of $\mathbf{q}(t_i)$ in $\mathcal{S}(\mathbf{q})$

From **Marsden & West 2001**

$$\begin{aligned} \mathcal{S}(\mathbf{q}) &= \lim_{n \to \infty} \sum_{i=0}^{n-1} L_{n,i}(\mathbf{q}) \quad \text{reducing } \mathbf{q} \text{ to } \mathbf{q}(t_0)\mathbf{q}(t_1) \cdots \mathbf{q}(t_n) \\ L_{n,i}(\mathbf{q}) &= h \cdot \mathcal{L}\Big(\frac{\mathbf{q}(t_i) + \mathbf{q}(t_{i+1})}{2}, \frac{\mathbf{q}(t_{i+1}) - \mathbf{q}(t_i)}{h}\Big) \quad \text{where } h = \frac{\hat{t}}{n} \text{ and } t_i = ih \end{aligned} \tag{4}$$

related to (2) as follows

$$\begin{aligned} \mathcal{S}(\mathbf{q}) &\approx -\log P(a_1 a_2 \cdots a_n) \quad \text{LHS of } (2) \\ L_{n,i}(\mathbf{q}) &\approx -\log P(a_{i+1} | q_i) \quad \text{from RHS of } (2) \\ &= K_i - V_i \quad \text{where } K_i = -\log P(q_i a_{i+1}) \quad \text{and} \quad V_i = -\log P(q_i) \end{aligned}$$

for 'kinetic minus potential energy' Lagrangian $\mathcal{L}(q, \dot{q}) = K(\dot{q}) - V(q)$.

But why replace Lagrangians $\mathcal{L}$ by probabilities P? Uncertainty from partiality.

David Bryant Mumford (born 11 June 1937) is an American mathematician known for his work in algebraic geometry and then for research into vision and pattern theory. He won the Fields Medal and was a MacArthur Fellow. In 2010 he was awarded the National Medal of Science.

David Mumford

David Mumford in 2010

Figure 1: From Wikipedia

David Mumford defines **Pattern Theory** as

> the analysis of the patterns generated by the world in any modality, with all their naturally occurring complexity and ambiguity, with the goal of **reconstructing the processes, objects and events** that produced them [1994, page 187]

where

> the pattern should not merely describe the 'pure' situation that underlies reality but the '**deformed**' situation that is actually **observed** in which the pure pattern may be hard to recognize. [2019, page 203]

A PROPOSAL: reconstruct processes, objects and events in paths (3) from some given initial state q_0 with*out* assuming $q_i = a_1 \cdots a_i$ and with cost equal to

$$\begin{aligned} -\log P(a_1 q_1 a_2 \cdots a_n q_n | q_0) &= \sum_{i=0}^{n-1} -\log P(a_{i+1} q_{i+1} | q_0 a_1 q_1 a_2 \cdots q_i) \\ &= \sum_{i=0}^{n-1} -\log P(a_{i+1} q_{i+1} | q_i) \quad \text{for many notions of } \to \end{aligned}$$

where in place of a Lagrangian, we have probabilities

$$P(a_i q_i | q_{i-1}) = P(a_i | q_{i-1}) P(q_i | q_{i-1} a_i)$$

that may arise, for example, from agent **policies** $P(a_i | q_{i-1})$ as well as state transitions $P(q_i | q_{i-1} a_i)$.

The world of the Lagrangian is just one corner in Ogden & Richards's Triangle of Meaning; another corner is language, which is linked to the world/physics by a third, agent/mind, observing both corners.

Without an agent, the language-world connection breaks. For example, Reichenbach's account of tense and aspect depends on a reference time R mediating between speech time S and event time E.

Meaning is built below around strings s that need *not* fully record each q_i or a_i from a path (3), strung out in pairs $(q_0, a_1)(q_1, a_2) \cdots (q_n, \emptyset)$.

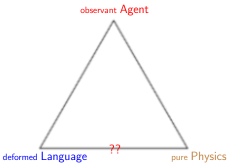


Figure 2: Triangle of Meaning

Step 2: Bounded granularities from finite-state nerve nets

The alphabet of strings below is derived from a Kleene nerve net spec $(k, m, s_1 \cdots s_m)$ describing

$$\begin{cases} k \text{ input cells} & \text{encoding } 2^k \text{ labels } a_i \\ m \text{ inner cells} & \text{encoding } \prod_{i=1}^{m} s_i \text{ states } q \end{cases}$$

which we augment over the set $\mathbb{R}$ of reals to a *sig* $\Sigma = (Q, A)$ consisting of a set A of k *acts* and a function $Q = \{(\mathbf{x}_i, Q(\mathbf{x}_i)) \mid 1 \leq i \leq m\}$ from m *variables* $\mathbf{x}_i$ to

$$\text{sets } Q(\mathbf{x}_i) \text{ of } s_i \text{ pairwise disjoint non-empty subsets of } \mathbb{R} \tag{5}$$

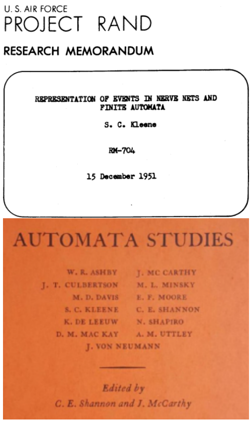
relative to which

- a Q-*state* is defined to be a function $q = \{(\mathbf{x}_i, q(\mathbf{x}_i)) \mid 1 \leq i \leq m\}$ choosing $q(\mathbf{x}_i) \in Q(\mathbf{x}_i)$,
- a (Q, A)-*box* is a pair (q, a) of a Q-state q and a set $a \subseteq A$ of acts,
- $\mathcal{B}(\Sigma)$ is the set of Σ-boxes, used below as the alphabet for strings s.

The requirement (5) on Q means we can approximate real numbers only to finite precision; e.g., for every integer $j > 0$ and $\mathbf{x} \in \mathrm{dom}(Q)$, we can let $Q(\mathbf{x})$ partition $\mathbb{R}$ into $2j^2 + 2$ intervals

$$(-\infty, -j), \quad [j, +\infty), \quad [-j + \frac{i}{j}, \ -j + \frac{i+1}{j}) \quad \text{for } i = 0, 1, \ldots, 2j^2 - 1$$

approximating reals in $[-j, j)$ to precision $1/j$ (cf. **blur**, Mumford 1994).



U. S. AIR FORCE
PROJECT RAND
RESEARCH MEMORANDUM

REPRESENTATION OF EVENTS IN NERVE NETS AND FINITE AUTOMATA
S. C. Kleene
RM-704
15 December 1951

AUTOMATA STUDIES

W. R. ASHBY
J. T. CULBERTSON
M. D. DAVIS
S. C. KLEENE
K. DE LEEUW
D. M. MAC KAY
J. MC CARTHY
M. L. MINSKY
E. F. MOORE
C. E. SHANNON
N. SHAPIRO
A. M. UTTLEY
J. VON NEUMANN

Edited by
C. E. Shannon and J. McCarthy

Figure 3: Regular expressions for nerve nets (Kleene 1951/1956)

Step 3: Refinements within an institution

A sig $\Sigma = (Q, A)$ is refined by $\Sigma' = (Q', A')$ as follows

$$\begin{aligned} Q \leq Q' &\iff \mathrm{dom}(Q) \subseteq \mathrm{dom}(Q') \text{ and } (\forall \mathbf{x} \in \mathrm{dom}(Q))(\forall v' \in Q'(\mathbf{x}))(\exists v \in Q(\mathbf{x}))\ v' \subseteq v \\ \Sigma \sqsubseteq \Sigma' &\iff Q \leq Q' \text{ and } A \subseteq A' \end{aligned}$$

and whenever $\Sigma \sqsubseteq \Sigma'$, we define

- the Q-*reduct of a* Q'-*state* q', written ${q'}_Q$, to be the Q-state mapping $\mathbf{x} \in \mathrm{dom}(Q)$ to the unique $v \in Q(\mathbf{x})$ such that $q'(\mathbf{x}) \subseteq v$,
- the Σ-*reduct of a* Σ'-*box* (q', a') to be the Σ-box $({q'}_Q, a' \cap A)$, and
- the Σ-*reduct of a string* $\mathrm{s}' = \alpha'_1 \cdots \alpha'_n$ of Σ'-boxes α'_i to be the string $\rho_\Sigma(\mathrm{s}) = \alpha_1 \cdots \alpha_n$ of Σ-boxes α_i equal to the Σ-reduct of α'_i for $1 \leq i \leq n$.

The map $\Sigma \mapsto \mathcal{B}(\Sigma)$ is a contravariant functor $\mathcal{B}$ from $\sqsubseteq$ on sigs if whenever $\Sigma \sqsubseteq \Sigma'$,

$$\mathcal{B}(\Sigma', \Sigma) \text{ maps a } \Sigma'\text{-box to its } \Sigma\text{-reduct.}$$

Similarly, $\Sigma \mapsto \mathcal{B}(\Sigma)^*$ is a contravariant functor **Mod** from $\sqsubseteq$ if whenever $\Sigma \sqsubseteq \Sigma'$,

$$\mathbf{Mod}(\Sigma', \Sigma) \text{ maps } \mathrm{s}' \in \mathbf{Mod}(\Sigma') \text{ to its } \Sigma\text{-reduct } \rho_\Sigma(\mathrm{s}').$$

For every sig Σ, let $\mathit{Sen}(\Sigma)$ be a set Σ-sentences φ denoting regular languages $[\![\varphi]\!] \subseteq \mathcal{B}(\Sigma)^*$ (e.g., regular expressions, finite automata, MSO-sentences).
Sen is a *co*variant functor from $\sqsubseteq$ if for $\Sigma \sqsubseteq \Sigma'$, $\mathit{Sen}(\Sigma, \Sigma')$ maps $\varphi \in \mathit{Sen}(\Sigma)$ to $\mathit{Sen}(\Sigma, \Sigma')(\varphi)$ such that

$$[\![\mathit{Sen}(\Sigma, \Sigma')(\varphi)]\!] = \{\mathrm{s}' \in \mathbf{Mod}(\Sigma') \mid \rho_\Sigma(\mathrm{s}') \in [\![\varphi]\!]\} \tag{6}$$

establishing the *Satisfaction condition* on *institutions* (Goguen & Burstall 1992).

Step 4: Compression and other deformations

For refinements (inverting reducts) to lengthen strings (cf. *time step* h in (4)), recall that a string s is a *factor* of a string s' if $\mathrm{s}' = u\mathrm{s}v$ for some strings u and v. Strings in $\mathcal{B}(\Sigma)^*$ are compressed by a function κ that eliminates factors of the form $(\emptyset, \emptyset)$ or $(q, \emptyset)(q, a)$ by mapping the null string to itself, and for $\alpha \in \mathcal{B}(\Sigma)$, $\mathrm{s} \in \mathbf{Mod}(\Sigma)$,

$$\kappa(\alpha\mathrm{s}) := \begin{cases} \kappa(\mathrm{s}) & \text{if } \alpha = (q, \emptyset) \text{ where } q = \emptyset \text{ or for some } a,\ (q, a) \text{ is the first symbol of s} \\ \alpha\,\kappa(\mathrm{s}) & \text{otherwise.} \end{cases}$$

N.B. $\kappa(\kappa(\mathrm{s})) = \kappa(\mathrm{s})$ and for any finite set $\mathcal{A}$, some finite-state transducer computes κ restricted to $\mathcal{A}^*$.

Fernando 2023 puts

$$\begin{aligned} \mathbf{Mod}^\kappa(\Sigma) &:= \{\kappa(\mathrm{s}) \mid \mathrm{s} \in \mathbf{Mod}(\Sigma)\} \quad \text{no factors } (\emptyset, \emptyset) \text{ or } (q, \emptyset)(q, a) \\ \mathbf{Mod}^\kappa(\Sigma', \Sigma)(\mathrm{s}') &:= \kappa(\rho_\Sigma(\mathrm{s}')) \qquad \kappa\text{-compress reducts} \end{aligned}$$

(cf. **domain warping** in Mumford 1994 and **it from bit** (Wheeler 1990), complicating amalgamation)

$$\begin{aligned} \mathit{Sen}^\kappa(\Sigma) &:= \mathit{Sen}(\Sigma) \\ \mathit{Sen}^\kappa(\Sigma, \Sigma')(\varphi) &:= \langle\Sigma\rangle\varphi \quad \text{where} \\ &[\![\langle\Sigma\rangle\varphi]\!] = \{\mathrm{s}' \in \mathbf{Mod}^\kappa(\Sigma') \mid \kappa(\rho_\Sigma(\mathrm{s}')) \in [\![\varphi]\!]\} \quad \text{injecting } \kappa \text{ into (6).} \end{aligned}$$

Given $\varphi_i \in \mathit{Sen}(\Sigma_i)$ and $\Sigma_i \sqsubseteq \Sigma$ for $i \in \{1, 2\}$,

$$[\![\langle\Sigma_1\rangle\varphi_1 \wedge \langle\Sigma_2\rangle\varphi_2]\!] = \{\mathrm{s} \in \mathbf{Mod}(\Sigma) \mid \kappa(\rho_{\Sigma_1}(\mathrm{s})) \in [\![\varphi_1]\!] \text{ and } \kappa(\rho_{\Sigma_2}(\mathrm{s})) \in [\![\varphi_2]\!]\}$$

(cf. **multi-scale superposition** in Mumford 1994)

applying finite-state methods, supported by **theorem provers** such as Mona, modulo the reals $\mathbb{R}$ and probabilities (with noise, the paradigmatic deformation).