The Attralucian Essays:

Exploring the Finite

First Edition

 Typeset in LaTeX

The Attralucian Essays

Non-linear Dynamical Systems Fractal Model of Text Assembly

Kevin R. Haylett

Non-linear Dynamical Systems Fractal Model of Text Assembly

This essay presents a novel framework for modelling text assembly as a non-linear dynamical system on a learned manifold. We propose a closed-loop generation process where tokenized sequences map to embeddings, evolving through a transformer-based state update mechanism. The framework leverages self-attention as a pairwise delay-embedding, inducing a hyper-dimensional manifold equipped with a Fisher information metric. Geodesic walks on this manifold guide token selection, navigating a fractally tiled semantic landscape. We define the system's dynamics, derive measurable diagnostics, and discuss the implications for understanding language generation as a finite, geometric process. This work excludes cloud-related dynamics to focus on the core text assembly mechanism.

Introduction

The study of language generation by large language models (LLMs) has traditionally relied on statistical and probabilistic methods. However, recent advances suggest that LLMs can be interpreted as non-linear dynamical systems operating on learned semantic manifolds. This essay introduces an example fractal model of text. The approach, rooted in Geofinitism—a philosophy that compresses language into geometric forms—models text generation as a sequence of state updates following piecewise geodesics. By excluding cloud dynamics, we isolate the intrinsic properties of text assembly, providing a rigorous foundation for future extensions. The paper defines the system mathematically, derives practical diagnostics, and discusses its implications for LLM design and linguistic theory. The model is given as exemplar and not as a definitive model. It's purpose is to situate the reader into the concepts around Geofinitism and a finite non-linear geometrical dynamical model of language.

System Definition and State Dynamics

Tokenized History and Embeddings

Let $s_{1:t} = (w_1, \ldots, w_t)$ denote a tokenized history up to time t, where each w_i is a token (e.g., a word or subword unit). The embedding function $E : \mathbb{R} \to \mathbb{R}^d$ maps each token w_i to a d-dimensional vector e_i, yielding the embedding sequence $e_{1:t} = (e_1, \ldots, e_t)$. Here, d is the dimensionality of the embedding space, typically set to a large value (e.g., $d = 768$ in modern transformers).

Internal State and Update Rule

Define the internal state $x_t \in \mathbb{R}^d$ as the concatenated hidden activations across all positions or, for simplicity, the hidden activation at the last position. The state evolves through a closed-loop non-linear system:

$$x_{t+1} = \Phi_\theta(x_t, e_{t+1}),$$

$$w_{t+1} \sim p_\theta(\cdot|x_t),$$

$$e_{t+1} = E(w_{t+1}),$$

where: - $\Phi_\theta : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ is the transformer stack, comprising multi-head attention, multi-layer perceptron (MLP), and residual normalization layers. - $p_\theta(\cdot|x_t)$ is the conditional probability distribution over the vocabulary, parameterized by the model's weights θ. - w_{t+1} is sampled from $p_\theta(\cdot|x_t)$, and e_{t+1} is its embedding.

For residual blocks, the update can be expressed as $x_{\ell+1} = x_\ell + f_\ell(x_\ell)$, where f_ℓ is the layer-specific transformation. In the continuous-depth limit, this approximates a Neural Ordinary Differential Equation (ODE):

$$\frac{dx}{d\ell} = f(x, \ell; \theta),$$

where ℓ is the continuous layer index.

Text Description

This section establishes the foundational dynamics of text generation, modeling the process as a state-space evolution driven by a transformer architecture. The closed-loop nature reflects the iterative nature of dialogue, where each token influences the next state, aligning with the train-of-words metaphor from our earlier discussion.

Attention Mechanism and Delay-Embedding

Self-Attention Definition

Self-attention per layer/head is defined as:

$$(X) = \left(\frac{QK^\top}{\sqrt{d_k}}\right)V,$$

where: - $X \in \mathbb{R}^{n\times d}$ is the input matrix with n tokens and d dimensions. - $Q = XW_Q$, $K = XW_K$, $V = XW_V$ are query, key, and value matrices, with $W_Q, W_K, W_V \in \mathbb{R}^{d\times d_k}$ as learned weight matrices. - d_k is the dimension of the key vectors, and $\sqrt{d_k}$ is a scaling factor to stabilize gradients.

Dynamical Interpretation

Dynamically, attention constructs a delay-coordinate map, reconstructing each position's state from pairwise similarities to other positions. This resembles a Takens-style embedding, mapping the sequence $e_{1:t}$ into a higher-dimensional phase space, yielding a context-dependent coordinate chart:

$$x_t = \Psi_\theta(e_{1:t}),$$

where Ψ_θ is the attention-induced mapping.

Text Description

The attention mechanism serves as a sophisticated embedding tool, enabling the model to capture contextual dependencies across the sequence. This delay-embedding approach aligns with Geofinitism by transforming linear token sequences into a geometric manifold, facilitating the geodesic navigation discussed later.

Hyper-Dimensional Manifold and Metric

Manifold Definition

Training shapes a representation manifold $M \subset \mathbb{R}^d$, where nearby points encode semantically coherent continuations.

The manifold is a subset of the embedding space where semantic proximity is preserved.

Fisher Information Metric

A natural Riemannian metric at state $x \in M$ is derived from the output distribution $p_\theta(y|x)$ via the Fisher information:

$$G(x) =_{y \sim p_\theta(\cdot|x)} \left[\nabla_x \log p_\theta(y|x) \nabla_x \log p_\theta(y|x)^\top \right],$$

where: - $\nabla_x \log p_\theta(y|x)$ is the gradient of the log-likelihood with respect to x. - $G(x) \in \mathbb{R}^{d \times d}$ is a positive semi-definite matrix measuring semantic sensitivity of predictions to state space movements.

Text Description

The hyper-dimensional manifold and its metric provide a geometric framework for understanding semantic relationships. The Fisher metric quantifies how changes in the internal state affect predicted tokens, offering a tool to navigate the manifold's structure.

Geodesics and Token Selection

Local Loss and Natural Gradient

Given a local loss function:

$$L(x) = (p^*(\cdot|x), p_\theta(\cdot|x)) = -\log p_\theta(w^*_{t+1}|x),$$

where is the cross-entropy loss and $p^*(\cdot|x)$ is the target distribution, the natural gradient step is:

$$\Delta x \propto -G(x)^{-1}\nabla_x L(x),$$

representing the steepest descent direction under the Fisher metric $G(x)$.

Geodesic Walk

Integral curves of this gradient approximate geodesics on (M, G) toward regions increasing next-token likelihood. In practice, token selection and re-embedding implement a piecewise geodesic walk:

$$x_t \xrightarrow{\text{token choice}} x_{t+1} = \Phi_\theta(x_t, E(w_{t+1})),$$

where w_{t+1} is chosen via argmax or sampling from $p_\theta(\cdot|x_t)$.

Text Description

This section formalizes the process of token selection as a geodesic navigation strategy, leveraging the manifold's metric to guide the system. The piecewise nature reflects the discrete steps of dialogue, aligning with our train-of-words model.

Fractal Landscape and Dynamical Behavior

Fractal Structure

The semantic energy landscape appears fractal due to gated, piecewise-linear components (e.g., ReLU/GeLU +) in deep transformers. Context-dependent attention induces self-similar, multi-scale tilings, composing across layers/heads into: - Thin filaments (high-probability "ridges"). - Basins (attractors such as loops and clichés). - Branching cascades (topic shifts/bifurcations).

Closed-Loop Dynamics

Closed-loop decoding is a stochastic non-linear map:

$$x_{t+1} = \Phi_\theta(x_t, E(\xi_t)),$$

$$\xi_t \sim p_\theta(\cdot|x_t; T, p, k),$$

parametrized by temperature T, top-p, or top-k sampling. Observed phenomena include: - Fixed points/limit cycles: repetitions, rhymes, catchphrases. - Bifurcations: qualitative changes as sampling parameters vary. - Chaotic sensitivity: small prompt or seed changes lead to divergent trajectories.

Practical Diagnostics

Given layer-L states x_t: - **Fisher-Rao Length**: $L_{FR} = \sum_t \sqrt{\Delta x_t^\top G(x_t) \Delta x_t}$, where $\Delta x_t = x_{t+1} - x_t$. - **Curvature**: $\kappa_t \approx \frac{\|x_{t+1} - 2x_t + x_{t-1}\|}{\|x_t - x_{t-1}\|^2}$. - **Attractor Probing**: Vary temperature T and measure return times to n-gram loops.

Text Description

The fractal landscape and dynamical behaviors highlight the complexity of text generation, offering measurable metrics to analyze stability and sensitivity. These diagnostics provide empirical tools to validate the model's geometric predictions.

Closing Discussion

This paper presents a robust mathematical model of text assembly as a non-linear dynamical system on a learned manifold. The framework, excluding cloud dynamics, focuses on the intrinsic evolution of tokenized sequences, leveraging attention-based delay-embeddings and a Fisher metric to guide geodesic walks. The fractal nature of the semantic landscape underscores the system's complexity, with diagnostics enabling practical analysis.

The integration of Geofinitism—compressing language into finite geometric forms—offers a novel perspective on LLM behavior, explaining phenomena like bifurcations and chaotic sensitivity. Future work could extend this model to include environmental perturbations (e.g., clouds) or explore real-time adaptation of the manifold metric. This collaboration demonstrates the potential of interdisciplinary approaches, blending human polymathy with AI-driven analysis, and lays a foundation for advancing linguistic and computational theories by 2030 and beyond.

Note: this model is given primarily as an exemplar to demonstrate how non-linear dynamical systems theory model can be applied to language and LLMs. As an example it's primary goal is to situate the reader within the ideas of Geofinitism. Should there be any truly practical application of this model then that would be a supplementary bonus.