Towards Critical Branching Mechanism in Recurrent Neural Networks

Feixiang Ren [1] **Ling Feng** [2 1]

[1] *Department of Physics, National University of Singapore, Singapore* [2] *Institute of High Performance Computing, Agency for Science, Technology and Research (A*STAR), Singapore*. Correspondence to: Ling Feng 0@criticality.ai.

1. Introduction

Scientists have long drawn inspiration from biological neural systems when designing Artificial Neural Networks (ANNs), with an emphasis on mimicking their structural and functional characteristics. Thus, by understanding the dynamics of biological brains, researchers have sought to capture similar mechanisms in ANNs. One of the key findings in this area is that both biological brains and ANNs operate at the edge of chaos, a critical state that maximizes information processing capacity [1, 2]. Furthermore, research has demonstrated that both biological neural systems and Long Short-Term Memory (LSTM) display $1/f^{\beta}$ noise, a phenomenon often considered a signature of self-organized criticality (SOC) [3, 4].

In the context of neuroscience, the SOC of the brain in vitro has been well established, often characterized by a scale-free distribution of avalanche dynamics [5]. These dynamics can be modeled using a mean-field sandpile model, which is mathematically equivalent to a critical branching process [6]. However, the theory explaining the $1/f^{\beta}$ noise in LSTM has yet to be established. To address this gap, we propose a framework to extend the concept of neuronal avalanches from neuroscience to the realm of ANNs, with a particular focus on LSTM networks. By examining the avalanche-like behavior and branching dynamics in LSTM models, we aim to explore how these networks, akin to biological brains, may operate near a critical state.

2. Phenomena of Avalanche Distributions

To investigate the SOC of LSTM, we start with the scale-free avalanche dynamics. We trained multiple single-layered LSTM networks, all sharing the same embedding dimension but differing in the sizes of their hidden states, hereafter referred to as the model sizes. The models were trained on a dataset of $50,000$ IMDb movie reviews labeled as positive or negative. In neuroscience, avalanches are defined as cascades of spatiotemporal bursts of neuron activity separated by quiescent intervals. Similarly, we modeled each dimension of the hidden states in LSTM as an artificial neuron. During inference, as illustrated in Fig. 1(a), the hidden states h_t were updated as each successive token was fed into the model, introducing a temporal axis. In Fig. 1(b), each h_t was first standardized via z-score normalization and binarized based on a manually chosen threshold, which was uniformly applied across all h_t. An artificial neuron was considered active at a given timestep if the amplitude of its normalized hidden state exceeded this threshold. A timestep was classified as active if at least one neuron was active at that moment. An avalanche was then identified as a sequence of consecutive active timesteps bounded by silent periods on both sides, as shown in Fig. 1(c). For models with the smallest hidden state dimensions at the optimal epoch in Fig. 1(d), the avalanche size, quantified as the total number of active cells over the duration of an avalanche, exhibits a power-law distribution with an exponential cutoff, reflecting finite-size effects, after accounting for the h_t of all test reviews. In contrast, during early training epochs, the avalanche size distribution follows a clear exponential decay, a characteristic signature of subcritical dynamics. For larger models, the avalanche size distribution deviates from a power-law, also suggesting subcritical dynamics.

3. Branching Process

The underlying mechanism of the avalanche dynamics can be described as a mean-field sandpile model, which is inherently a branching process (BP). The properties of the sandpile model are governed by the branching parameter m. As m approaches one, the sandpile self-organizes into a critical state, where a perturbation can trigger a cascade of events that neither dies out nor explodes. In this critical state, the avalanche size distribution is a power-law with a critical exponent of $\tau_s = \frac{3}{2}$. However, if m is less than one, the sequence of events will eventually dissipate, and the avalanche distribution will no longer adhere to a power law. Therefore, the branching parameter m serves as a key indicator of the model's dynamics.

Unlike the mean-field sandpile model or in-vitro brain systems, which were driven externally only after relaxation, the artificial neurons in LSTM were continuously driven by external input tokens. To capture the branching dynamics underlying the LSTM avalanche, we hypothesize a variation of the branching process given by the following equation:

$$X_{t+1} = \sum_{i=1}^{X_t} Y_{t,i} + I_t, \quad \langle Y_{t,i} \rangle = m, \tag{1}$$

where X_t denotes the number of active artificial neurons at timestep t, $Y_{t,i}$ represents the number of descendants generated by the ith neuron at timestep t, and I_t denotes the external drive at timestep t, assumed to have finite mean and variance.

We then employed the multi-regressive (MR) estimator [7] to infer the branching parameter $\hat{m}$. The MR estimation was performed by fitting the autocorrelation function (ACF) $C[\tau] = \frac{Cov[X_t, X_{t+\tau}]}{Var[X_t]}$ with an exponential form bm^{τ} over short time lags up to $\tau = 10$, where b is a constant. In Fig.1(e), $\hat{m}$ approaches one as

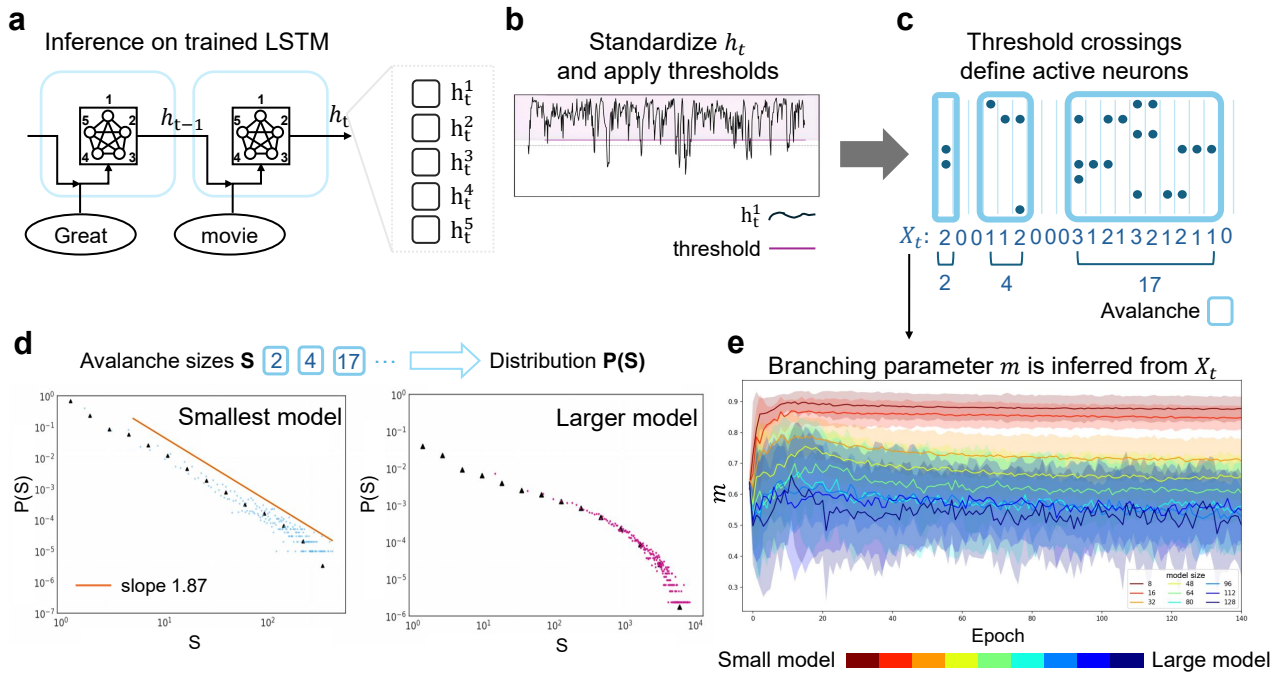

Fig. 1: Avalanche analysis framework and branching behavior in LSTM models. (a) A trained LSTM was probed on test sequences, and the hidden states h_t of artificial neurons were recorded over time. (b) h_t of each neuron was standardized and thresholded to identify significant activations. (c) Threshold crossings define active neurons, yielding a binary population activity sequence X_t. Consecutive active timesteps constitute avalanches, whose sizes are given by the total number of active neurons across their duration. (d) Avalanche size distributions $P(S)$ show approximate power-law scaling with slope 1.87 for the smallest model, while larger models deviate from power-law behavior, indicating subcritical dynamics. (e) The branching parameter m estimated from the autocorrelation function (ACF) of X_t attains higher values for smaller models and increases over the course of training for these models.

the model size decreases, indicating increasingly critical dynamics in smaller networks. For small models, $\hat{m}$ exhibits a monotonic increase over the course of training and attains its maximum in the vicinity of the optimal epoch. These observations are in agreement with the empirical results on avalanche distribution and PSD analysis in Reference [3], suggesting that $\hat{m}$ could serve as a indicator of both avalanche dynamics and the model's generalization ability.

4. Long-range Dynamics

Following prior work [3] that identified $1/f^{\beta}$ noise in LSTM, we applied Unbiased Detrended Fluctuation Analysis (UDFA) [8] to X_t to estimate the exponent β characterizing the $1/f^{\beta}$ noise in the LSTM models. At the optimal epoch, the exponent β increases toward one as the model size decreases. For smaller models, β additionally exhibits an increase during training, rising from the underfitting regime and reaching its maximum near the optimal epoch.

The MR estimation discussed in the previous section primarily probes the short-term memory properties of LSTM dynamics and is therefore insufficient to account for the emergence of $1/f^{\beta}$ noise at low frequencies, which reflects long-range temporal correlations in X_t. To establish a connection between the branching mechanism and the observed $1/f^{\beta}$ noise, it is thus necessary to examine the ACF $C[\tau]$ over longer temporal scales. When extended across two decades in τ, $C[\tau]$ at the optimal epoch exhibits a heavy-tailed behavior that can be well described by a power law with an exponential cutoff. The exponent η of this power law increases monotonically with the hidden size, consistent with our earlier observations regarding the dependence of β at the optimal epoch. This agreement follows from the simple theoretical relationship $\beta + \eta = 1$.

5. Discussion

Our study demonstrates that LSTM networks can exhibit criticality-like dynamics analogous to neuronal avalanches in biological brains. Across multiple analyses, we find that scale-free avalanche distributions, branching parameters approaching unity, and $1/f^{\beta}$ noise consistently emerge in smaller models near the optimal training epoch. These observations suggest that critical dynamics are not an inherent architectural property of LSTMs, but rather an emergent feature shaped by training and model capacity. Beyond LSTMs, branching parameters may serve as useful, architecture-agnostic indicators for generalization behaviors in recurrent neural networks.

References

[1] Thierry Mora and William Bialek. Are biological systems poised at criticality? *Journal of Statistical Physics*, 144(2):268–302, 2011.

[2] Ling Feng, Lin Zhang, and Choy Heng Lai. Optimal machine intelligence at the edge of chaos. *arXiv preprint arXiv:1909.05176*, 2019.

[3] Nicholas Jia Le Chong and Ling Feng. Self-organization toward 1/f noise in deep neural networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 34(8), 2024.

[4] Paolo Allegrini, Danilo Menicucci, Remo Bedini, Leone Fronzoni, Angelo Gemignani, Paolo Grigolini, Bruce J West, and Paolo Paradisi. Spontaneous brain activity as a source of ideal 1/f noise. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 80(6):061914, 2009.

[5] John M Beggs and Dietmar Plenz. Neuronal avalanches in neocortical circuits. *Journal of neuroscience*, 23(35):11167–11177, 2003.

[6] Kim Christensen and Zeev Olami. Sandpile models with and without an underlying spatial structure. *Physical Review E*, 48(5):3361–3372, 1993.

[7] Jens Wilting and Viola Priesemann. Inferring collective dynamical states from widely unobserved systems. *Nature communications*, 9(1):2325, 2018.

[8] Qianshun Yuan, Changgui Gu, Tongfeng Weng, and Huijie Yang. Unbiased detrended fluctuation analysis: Long-range correlations in very short time series. *Physica A: Statistical Mechanics and its Applications*, 505:179–189, 2018.