Towards Critical Branching Mechanism in Recurrent Neural Networks

Feixiang Ren[1] and Feng Ling[2]

[1]Department of Physics, National University of Singapore, Singapore
[2]Institute of High Performance Computing, Agency for Science, Technology and Research (A*STAR), Singapore

Keywords: Self-organization, Branching process, Avalanche dynamics, Scale-free distribution, Phase transition

Extended Abstract

Motivation. Biological neural systems and artificial neural networks (ANNs) have been compared due to their structural and functional similarities. From the perspective of complexity science, both are hypothesized to operate near the edge of chaos, a critical state that maximizes information processing capacity [6, 5]. Furthermore, $1/f^{\beta}$ noise, commonly regarded as a hallmark of self-organized criticality (SOC), has been reported in both neural activities and Long Short-term Memory (LSTM) networks [1, 3]. In biological brains, SOC is well established and can be described by a mean-field sandpile model, which is mathematically equivalent to a critical branching process [4]. This criticality manifests as neuronal avalanches, namely cascades of activity whose size distributions are scale-free. In contrast, a branching-process interpretation of the dynamical signatures is still lacking for LSTMs. Therefore, this work aims to examine the avalanche-like behaviors and their underlying branching dynamics in LSTMs, to understand how these networks, akin to biological brains, operate near a critical state.

Approach and Methodology. We investigate single-layer LSTMs trained on IMDb sentiment classification, systematically varying the hidden-state dimension. Inspired by neuronal avalanches [2], each component of the hidden state is treated as an artificial neuron. During inference, as input tokens are sequentially processed, the resulting temporal signals are normalized to ensure comparability across neurons and subsequently thresholded to yield binary activation patterns, from which a population activity signal X_t is constructed as the number of active neurons at each timestep. Avalanches are then defined as contiguous sequences of active timesteps separated by periods of inactivity, with the avalanche size S quantified as the total number of activations accumulated within each such sequence. To probe the propagation mechanism underlying these avalanches, we model the dynamics as a driven branching process,

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

where $Y_{t,i}$ represents the number of descendants generated by the i^{th} neuron at timestep t, m is the branching parameter and I_t represents the continuous external drive induced

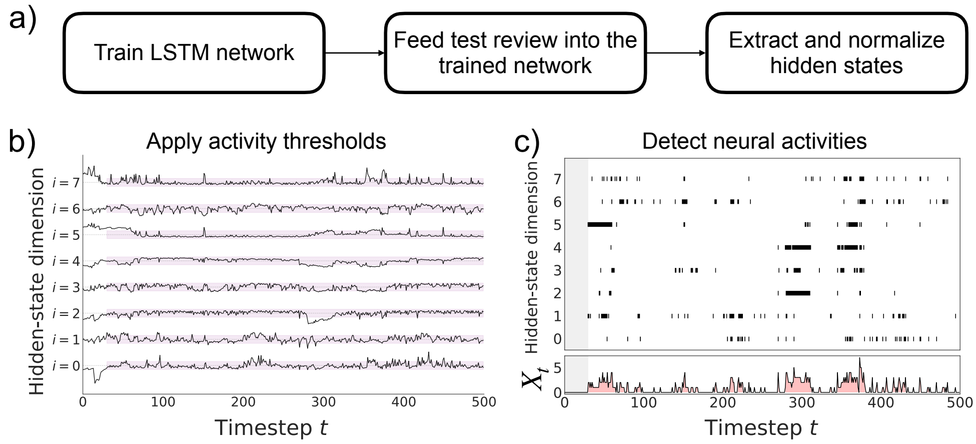

Figure 1: Extraction of neural activities from an LSTM network. (a) Schematic of the analysis pipeline. LSTM processes test reviews to generate hidden states $\boldsymbol{h}_t$, which are recorded and normalized for subsequent analysis. (b) Time series of normalized hidden states for selected dimensions i. A uniform activity threshold (shaded regions) is applied independently to each normalized hidden-state dimension to identify significant activation events. For illustration, the hidden-state dimensionality is set to eight; in the experiments, networks with varying hidden dimensions were analyzed. (c) Binary raster plot of detected neural activities across dimensions after thresholding (top), together with the corresponding X_t, defined as the number of active neurons at each timestep t (bottom). Consecutive timesteps for which $X_t \neq 0$ (red regions) are identified as avalanches.

by input tokens. We estimate m from the autocorrelation function of X_t using the multi-regressive (MR) estimator at different training stages of LSTMs with different hidden-state dimensionality [7].

Results. Two consistent trends are observed. First, the smallest models near their optimal training epoch exhibit avalanche-size distributions that approximate a power law with an exponential cutoff, whereas early training stages and larger models show predominantly exponential decay, indicative of subcritical dynamics. This suggests a transition toward near-critical behavior during training in smaller models. Second, the inferred branching parameter increases over the course of training and approaches unity most closely in these smaller models, coinciding with the regime in which scale-free avalanche statistics are most broadly distributed. In contrast, larger models maintain consistently lower branching parameters throughout training.

Conclusions and Outlook. Taken together, these findings indicate that scale-free avalanche statistics and near-critical branching in LSTMs appear to be an emergent dynamical regime shaped by both training stage and hidden-state capacity, rather than a fixed architectural property. Smaller models are more prone to self-organize toward a near-critical branching state, suggesting that the branching parameter may serve as an architecture-agnostic indicator of model capacity and generalization. Ongoing work aims to formalize the connections between this branching-process framework and the $1/f^{\beta}$ noise observed in LSTMs, and to investigate whether similar phenomenology extends to broader classes of artificial neural network models.

References

[1] 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.

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

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

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

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

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