

Distributed Language Models with RNNs

Ting-Po Lee

Taman Narayan

CME 323, Spring 2016

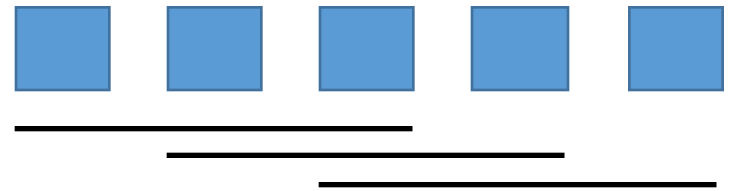
Background on Language Models and RNNs

- Goal: predict the next word/character
- RNNs vs. Markov models
- Loss function: $J = -\frac{1}{T} \sum_{t=1}^T \log(p_t^*)$
- Standard approach: sequential pass through data

RNNs in Distributed Context

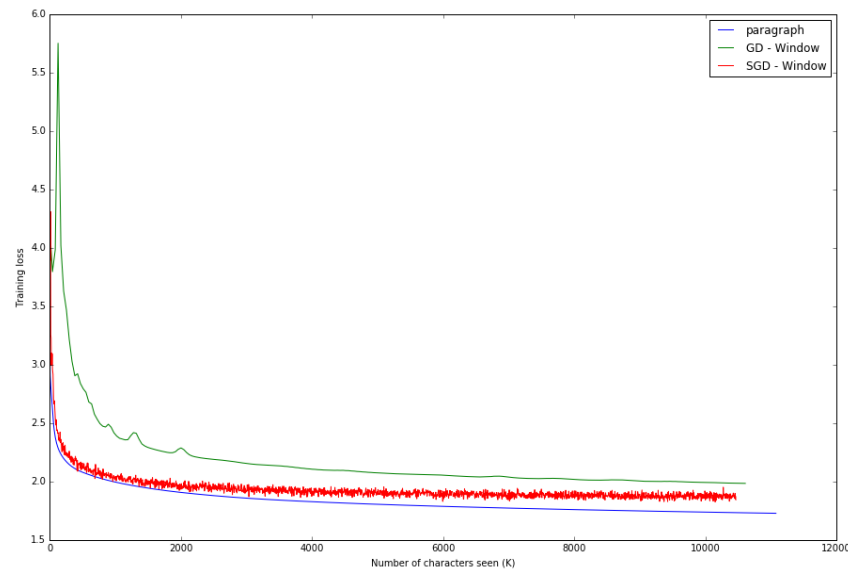
- Assume corpus stored in HDFS – how to distribute training?

- Idea 1: Parallel train on paragraphs
- Idea 2: Parallel train on sliding windows
- Idea 3: Sample from sliding windows



- We build ground-up base Spark implementation (plus dash of MLlib)
- Communication costs limited to AllReduce during parameter updates

Sample Results



Successfully reproducing a sentence from the corpus: And yet the bullshit you choose may be harder to eliminate than the bullshit that's forced on you.

Trying to uniquely combine parts of corpus: Otherwise these people are litably forced on you, the bullshit that sneaks into your life by tricking you is no one's fault