REINVENT THE WHEEL - GRU

Svätopluk Hanzel

FACULTY OF INFORMATION TECHNOLOGY

## 1 PROBLEM DESCRIPTION AND INTRODUCTION

A Gated recurrent unit (GRU) is a variant of RNN architecture for recurrent networks. It uses gating mechanisms to control the flow of information between its nodes.

This helps retaining long-term dependencies and combat the "short-term memory" issue of vanilla RNNs.

GRU networks outperform LSTM networks thanks to smaller number of parameters

GRU cell
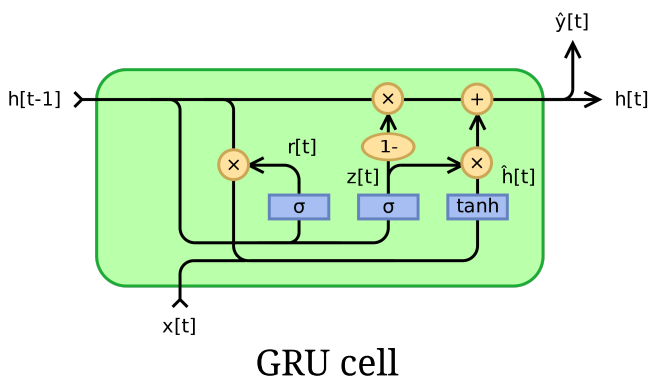
2 Implementation

This project implements a simple character-predicting GRU network. It is implemented in C++ with no external dependencies.

Most hyperparameters as well as training data can be changed using command-line options.

```
class Matrix
      - mathematical operations on matrices
class GRU
      - the main network class
      - uses Xavier initialization for its parameters
      - implements forward and backward passes
```

3 Training and evaluation

The project's functionality can be demonstrated by running the training on a sample dataset. For this purpose the Dinosaur list dataset was chosen, which contains **1143** unique names.

Training details:
- sequence of 20 characters
- 100 epochs
- learning rate 0.001
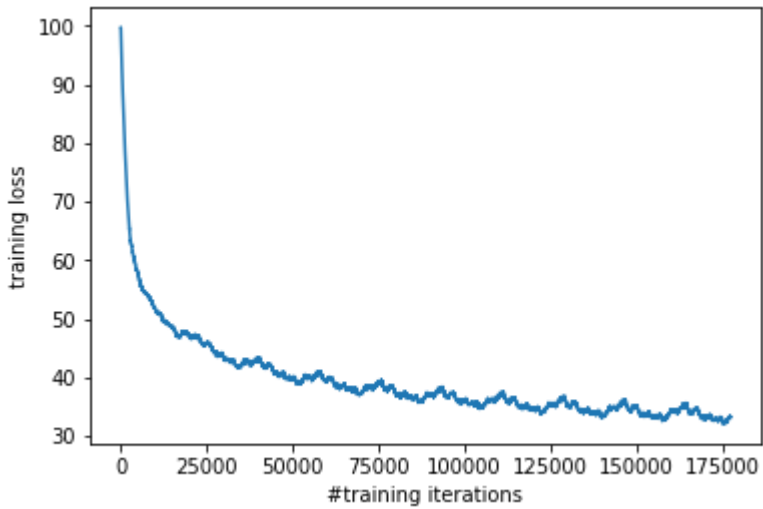
4 Results

The network is fully capable of learning and generating valid results:

```
    austrosaurus utarapateks rapator
venasugngowx sinornitholog
aladromeus toriosaurus
zenimaceitisaurus varaikenoplota
colopteryx velocirapteros
omaisaurus steronthodesceus
zhenyurocoese
```

As we can see, it learned to properly alternate vowels and consonants, and use suffixes such as -saurus, -us, or -raptor.

Resources
• https://colah.github.io/posts/2015-08-Understanding-LSTMs
• https://en.wikipedia.org/wiki/Gated_recurrent_unit
• https://d2l.ai/chapter_recurrent-modern/gru.html
• https://www.kaggle.com/kumazaki98/dinosaur-list