

A Machine Learning Perspective on Predictive Coding with PAQ by Knoll & de Freitas

Presentation by Phil Trommer

November 27, 2019

Overview

- 1 Introduction to PAQ
- 2 PAQ8L
 - Architecture
 - Neural Network
 - Model Mixer
 - Mixture of Experts
 - Updating & Filtering
- 3 Applications for PAQ8
- 4 References

Introduction to PAQ

- 1 Introduction to PAQ
- 2 PAQ8L
- 3 Applications for PAQ8
- 4 References

Introduction to PAQ

What is PAQ8

- What is it?
- How does it work?
- What makes it so famous?

What is PAQ?

- A lossless, open-source compression algorithm
- Brings high performance at the cost of increased memory usage and time consumption
- Related to PPM, is envisioned as PPMs improvement

Principles of PAQ

- Modeling combined with adaptive arithmetic encoding
- Open to additions and improvements
- Improves performance of PPM by including several predictors (i.e. models of data)
- Combines the result of the predictors

Introduction to PAQ

Exemplary Predictors

The order- n context predictor

- Examines the last n bits and counts the 1's and 0's
- Estimates probability whether next bit is 1 or 0 like PPM

Introduction to PAQ

Exemplary Predictors

The order- n context predictor

- Examines the last n bits and counts the 1's and 0's
- Estimates probability whether next bit is 1 or 0 like PPM

The sparse context predictor

- Context consists of a specific amount of non-contiguous bytes before the current bit
- Useful for some binary files

Introduction to PAQ

Exemplary Predictors

The order- n context predictor

- Examines the last n bits and counts the 1's and 0's
- Estimates probability whether next bit is 1 or 0 like PPM

The sparse context predictor

- Context consists of a specific amount of non-contiguous bytes before the current bit
- Useful for some binary files

Whole word order- n context

- Context is the latest n whole words
- Non-alphabetical characters are ignored and upper- or lower case letters are viewed as the same
- Very useful for text files

PAQ & Predictors

- PAQ encoder looks at the beginning of input file for deciding which predictors are used
- Ways to combine predictions change through with the different versions
- Each predictor outputs a pair of bit counts (n_0, n_1)
- Counts of each predictor are weighted with context length
- Those counts get summed up

1 Introduction to PAQ

2 PAQ8L

- Architecture
- Neural Network
- Model Mixer
- Mixture of Experts
- Updating & Filtering

3 Applications for PAQ8

4 References

PAQ8 - What's new?

- Predictors don't produce a pair of bit counts anymore
 \hookrightarrow those counts get weighted and normalized into the interval $[0, 1] \subset \mathbb{R}$
- Instead each predictor already outputs a probability
- *paq8l* is a stable version of paq8, released by Matt Mahoney

PAQ8 - What's new?

- Predictors don't produce a pair of bit counts anymore
 \hookrightarrow those counts get weighted and normalized into the interval $[0, 1] \subset \mathbb{R}$
- Instead each predictor already outputs a probability
- *paq8l* is a stable version of paq8, released by Matt Mahoney

PAQ8L - Machine Learning Perspective

- paq8l is the version of PAQ used by *Byron Knoll & Nando de Freitas*
- They try to show the possibilities of PAQ beyond data compression

Architecture of PAQ8

- Uses weighted combination of predictions from Large number of models
- Allows no-contiguous context matches
- paq8l uses **552** prediction models
- Combines the output of them into a single one
 - ↪ Passes this through an *adaptive probability map* (APM) before using the arithmetic coder

Architecture of PAQ8

- Uses weighted combination of predictions from Large number of models
- Allows no-contiguous context matches
- paq8l uses **552** prediction models
- Combines the output of them into a single one
↳ Passes this through an *adaptive probability map* (APM) before using the arithmetic coder

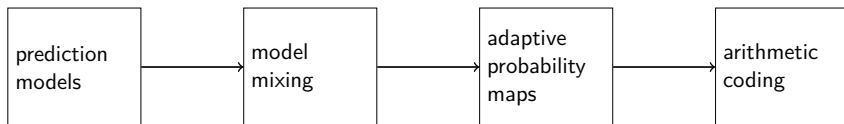


Figure: PAQ8 Architecture

Neural network

Neurons of a neural network

A **neuron** takes one or more **inputs** and gives an **output**.

Within the topic of machine learning, the neuron can be understood as a **function**.

Neural network

Neurons of a neural network

A **neuron** takes one or more **inputs** and gives an **output**.

Within the topic of machine learning, the neuron can be understood as a **function**.

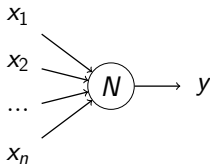


Figure: Neural network architecture

Neural network

Layer in neural network

A layer is a group of neurons.

Neural network

Layer in neural network

A layer is a group of neurons.

A neural network

Neural networks is defined by its layers:

- 1 input layer with n inputs
- 1 output layer with k outputs
- M layers between input and output layer (i.e. hidden layers)
- Layers can consist of different amounts of neurons

Neural network

Layer in neural network

A layer is a group of neurons.

A neural network

Neural networks is defined by its layers:

- 1 input layer with n inputs
- 1 output layer with k outputs
- M layers between input and output layer (i.e. hidden layers)
- Layers can consist of different amounts of neurons

General structure of neural network

Let it be an generic neural network with:

- x_1, \dots, x_n inputs and y_1, \dots, y_k outputs
- There are M different layers between input and output

Neural network

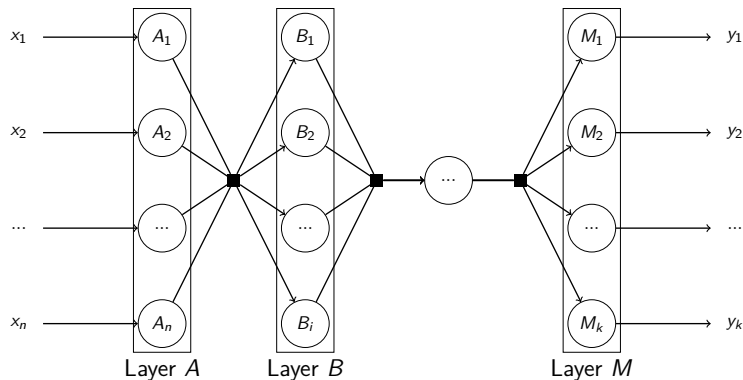


Figure: Neural network architecture

Model Mixer of paq8l

- Resembles a neural network with one hidden layer
- Hidden layer is between input and output layer
 - ↔ Artificial neurons take a set of weighted inputs
- Output is produced through activation function

Model Mixer of paq8l

- Resembles a neural network with one hidden layer
- Hidden layer is between input and output layer
↔ Artificial neurons take a set of weighted inputs
Output is produced through activation function

Differences between paq8l and neural networks

- 1 Weights for first and second layers are learned online and independently for all nodes:
 - Each node trained separately
 - reduces predictive cross-entropy error (unlike back propagation)
- 2 Hidden nodes are partitioned into seven sets

Applications for PAQ8

- 1 Introduction to PAQ
- 2 PAQ8L
- 3 Applications for PAQ8**
- 4 References

- 1 Introduction to PAQ
- 2 PAQ8L
- 3 Applications for PAQ8
- 4 **References**