Cache Miss Rate Predictability via Neural Networks

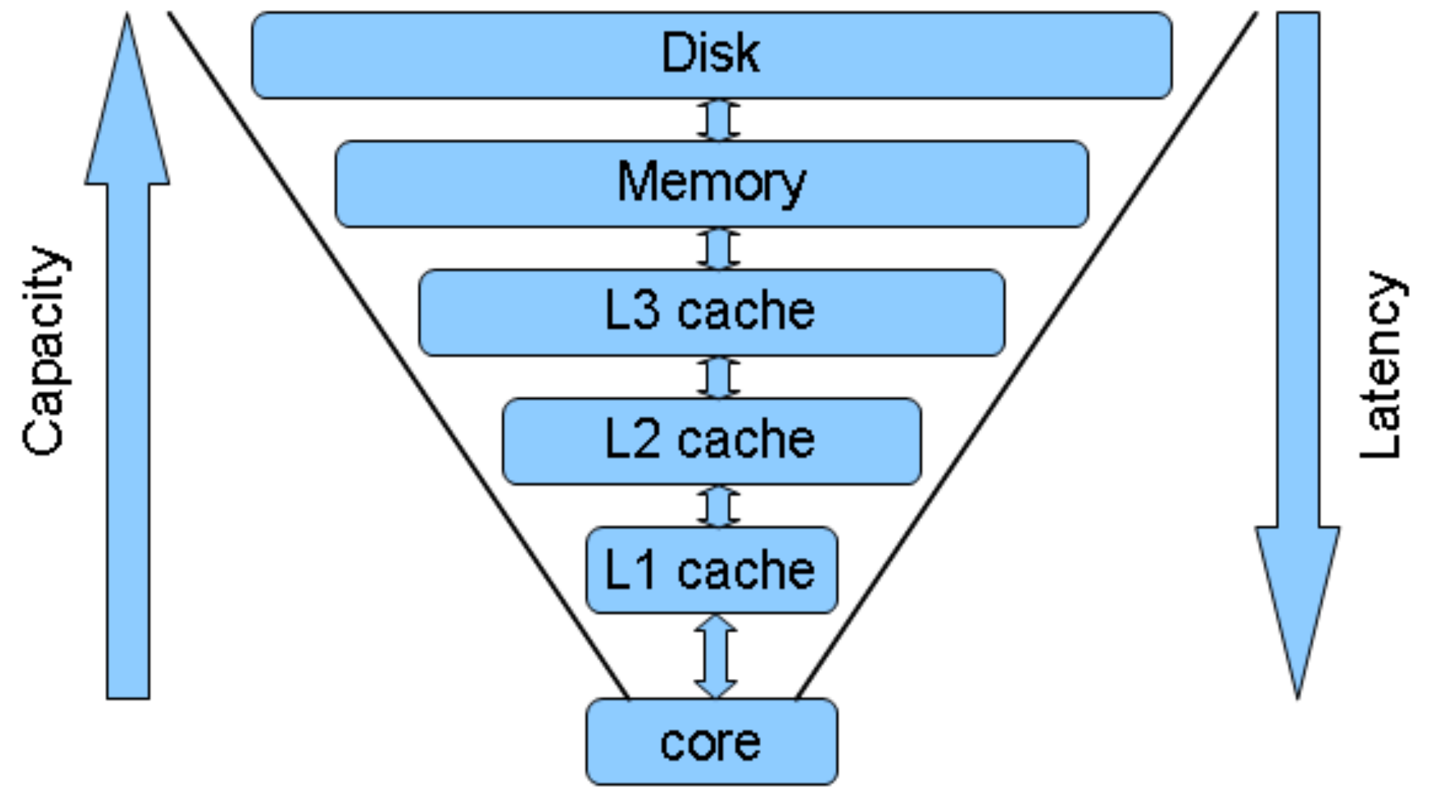
Rishikesh Jha, Arjun Kuravally, Saket Tiwari, Eliot Moss

UMASS AMHERST

Motivation

Can programs be characterized in part by its memory access patterns?

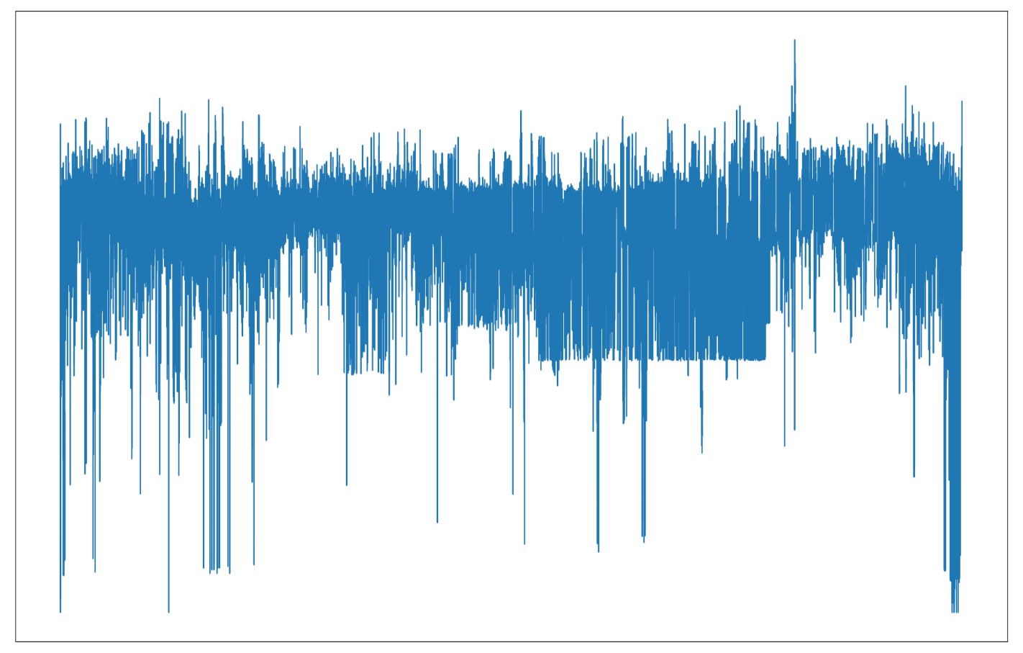
How predictable is a program's cache miss rate from interval to interval as it executes?



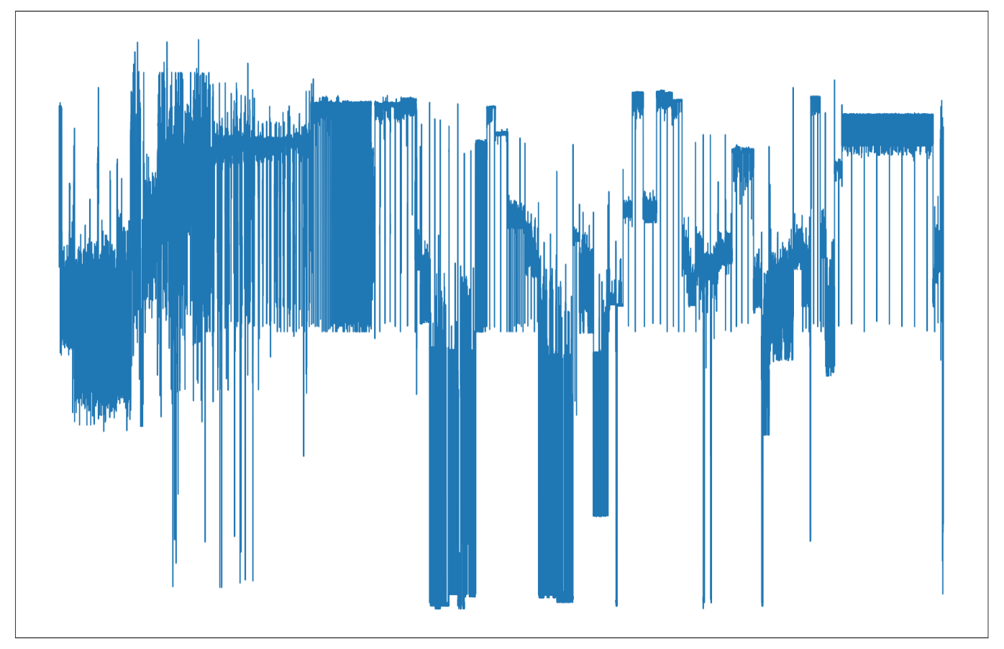
Memory Hierarchy

Dataset

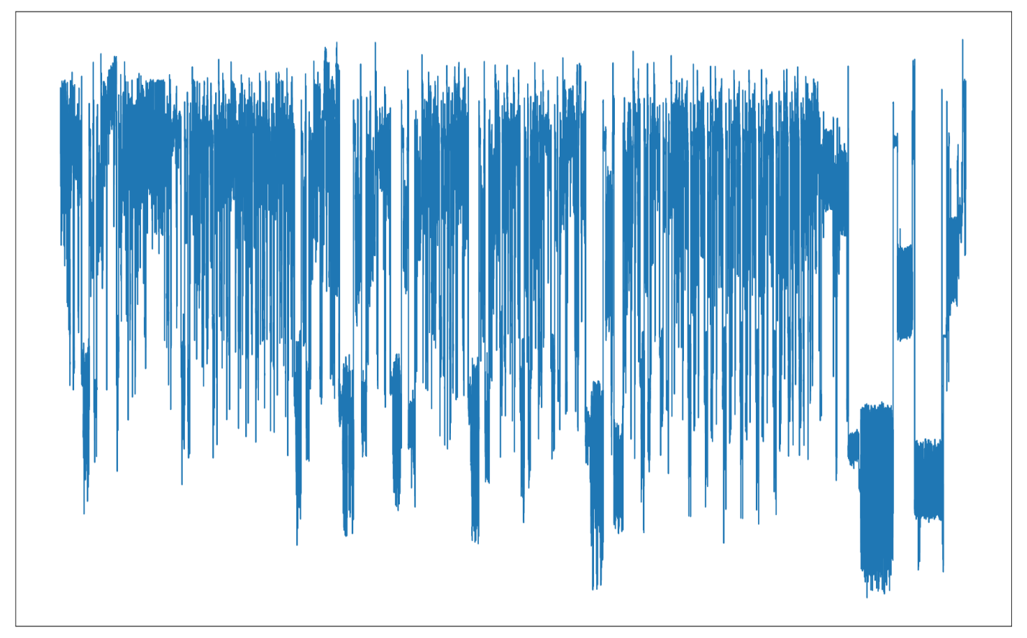
- Collected traces of every memory access made by the programs in *valgrind*.
- Ran Java programs on three different virtual machines
- Applied the least recently used (LRU) stack algorithm to obtain miss rates for various cache sizes.
- Rates are aggregated over 1M instructions/data accesses.



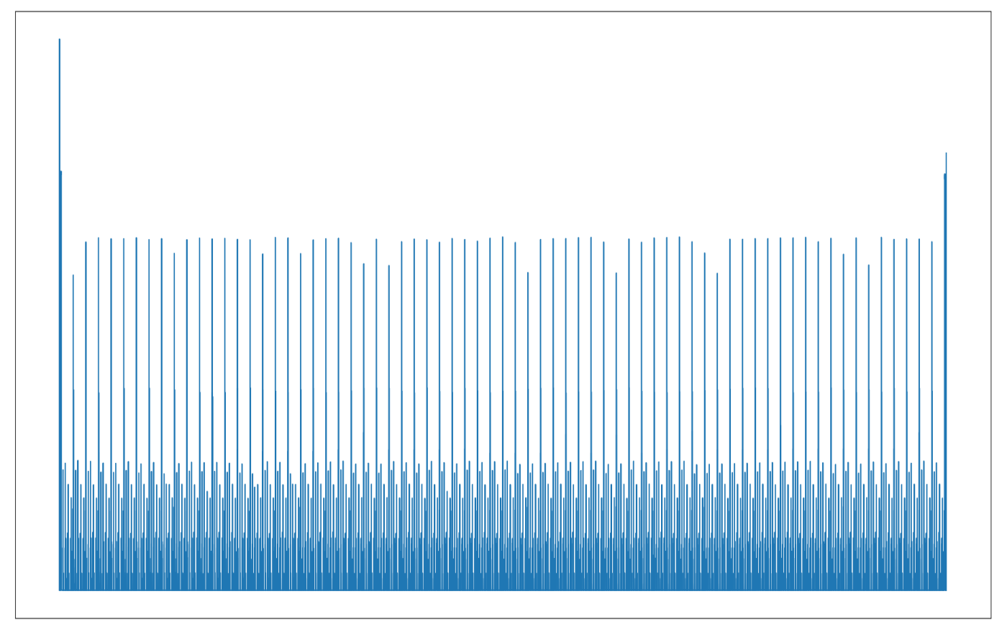
J9



Hotspot



C program



Fortran

Sequence Learning Models

We take inspiration from existing sequence learning tasks is NLP and audio waveform generation. We apply the following models to our problem.

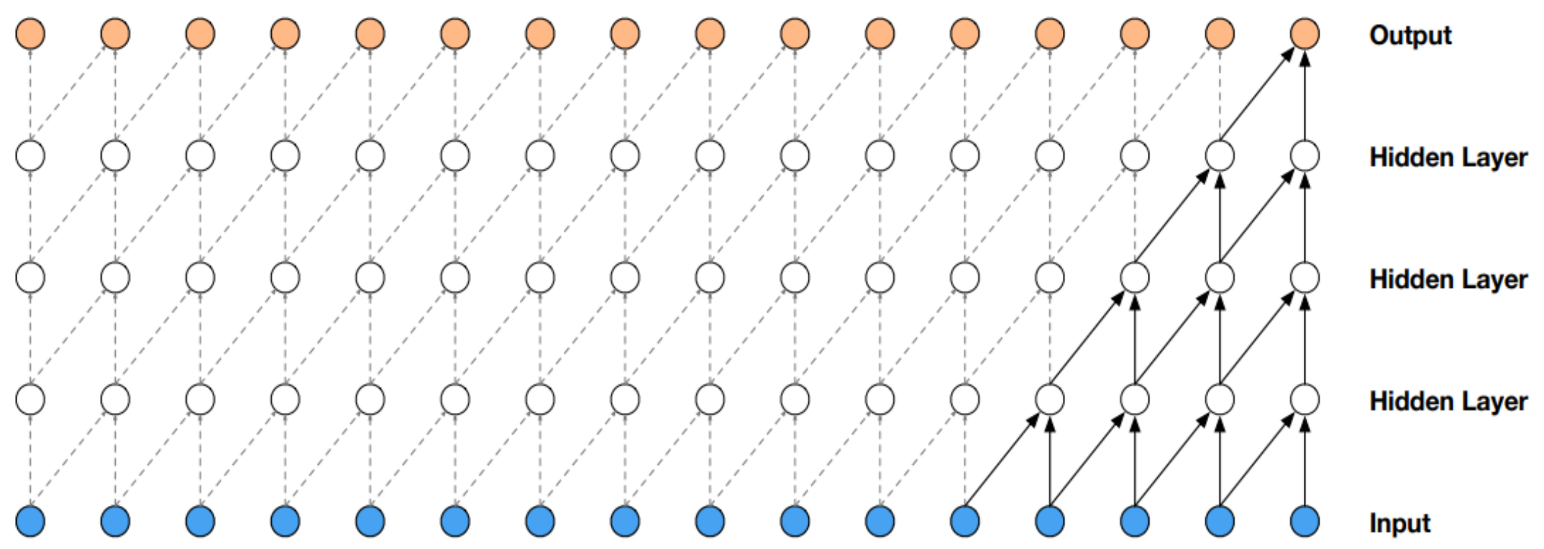
LSTM

We model the following equation using LSTMs:

$$\Pr(\mathbf{x}) = \Pi_{t=1}^{T} \Pr(x_t | x_1, ...x_{t-1}).$$

where x is the sequence of cache miss rates

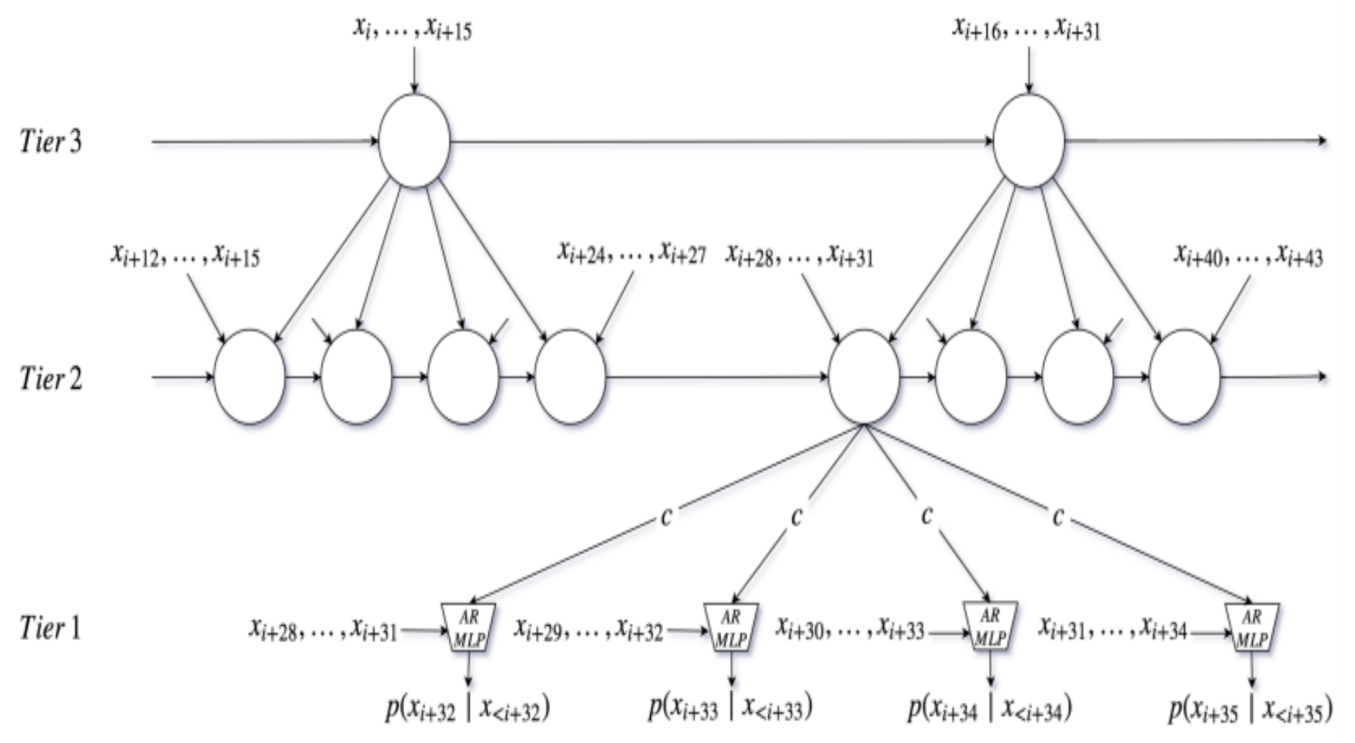
Wavenet



Wavenet model

$$\Pr(\mathbf{x}|\mathbf{h}) = \Pi_{t=1}^{T} \Pr(x_t | x_1, ..., x_{t-1}, \mathbf{h}).$$

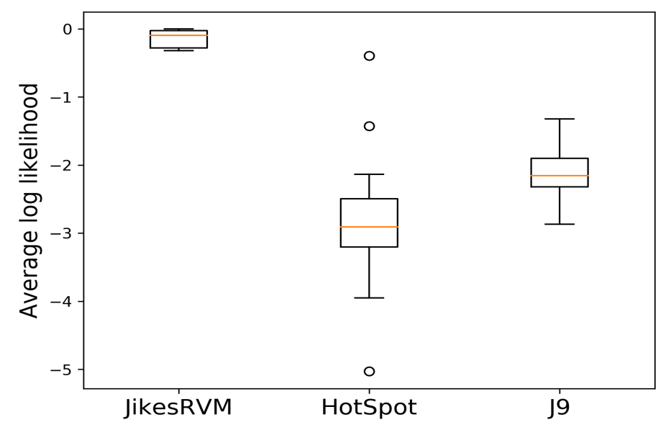
SampleRNN


$x_i, \dots, x_{i+15}$
$x_{i+16}, \dots, x_{i+31}$
Tier 3
$x_{i+12}, \dots, x_{i+15}$
$x_{i+24}, \dots, x_{i+27}$
$x_{i+28}, \dots, x_{i+31}$
$x_{i+40}, \dots, x_{i+43}$
Tier 2
c
c
c
c
Tier 1
$x_{i+28}, \dots, x_{i+31}$
AR MLP
$x_{i+29}, \dots, x_{i+32}$
AR MLP
$x_{i+30}, \dots, x_{i+33}$
AR MLP
$x_{i+31}, \dots, x_{i+34}$
AR MLP
$p(x_{i+32} | x_{<i+32})$
$p(x_{i+33} | x_{<i+33})$
$p(x_{i+34} | x_{<i+34})$
$p(x_{i+35} | x_{<i+35})$

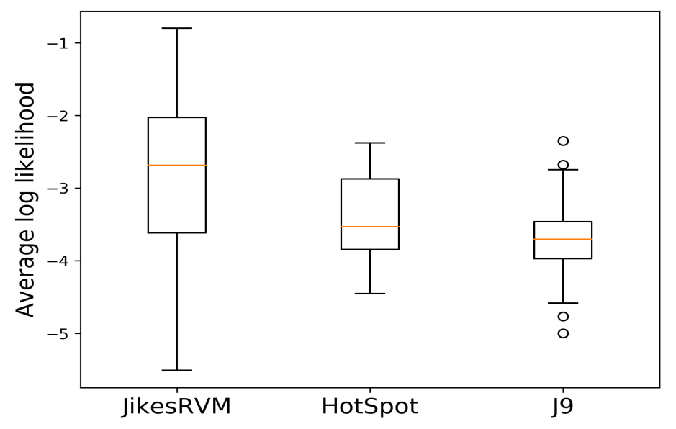
Sample RNN Model

Evaluation

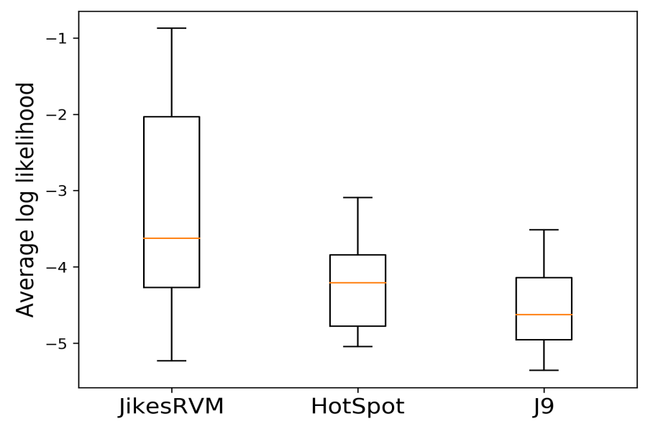
We evaluated our model on 227 different memory access traces.



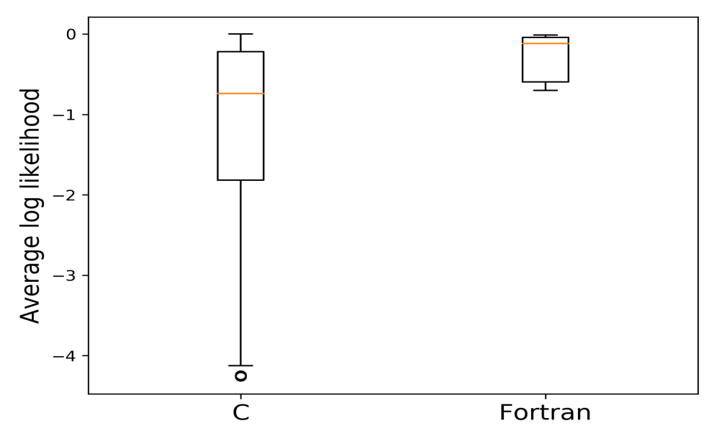
LSTM



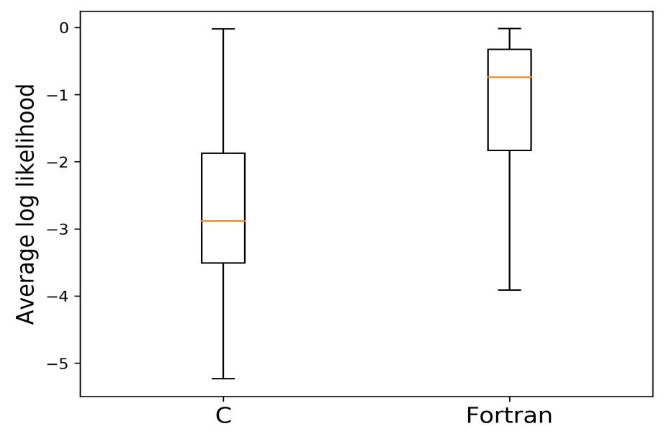
Wavenet



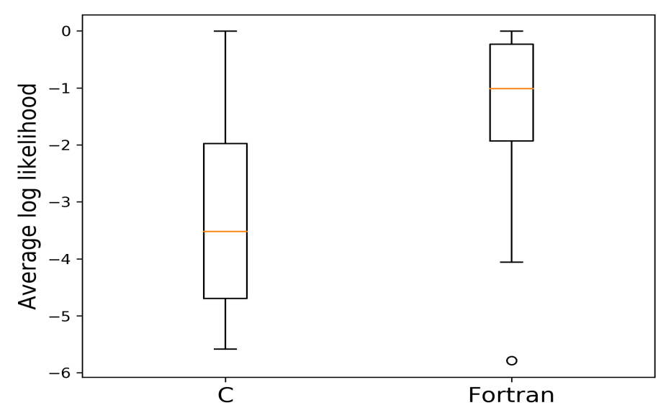
SampleRNN



LSTM



Wavenet



SampleRNN

Conclusions

- ➢ Introduced a new data set for analyzing the predictability of cache miss rates in program runs.
- ➢ Formulated a sequence modelling task to analyze predictability using Log Likelihood as a metric

Future Work

- ➢ Analyze factors that affect predictability.
- ➢ Forecasting cache miss rate for effective allocation of cache memory to programs