

January 15<sup>th</sup>, 2019

## Perceptron Learning for Cache Management

**Elvira Teran**

Texas A&M International University, USA

### Abstract

Reuse predictors predict whether a block will be used again before it is evicted from the cache. We propose to use several features associated to an access as input to our predictors, the goal is to correlate past behavior with future accesses. The higher accuracy obtained by our predictors yield superior overall performance.



This talk will present two techniques that apply perceptron learning to cache management. For our Perceptron Predictor we propose using perceptrons to predict cache line reuse. Perceptron trains weights selected by hashes of multiple features, including the PC of the memory access instruction, other recent PCs, and two different shifts of the tag of the referenced block.

The Multiperspective Reuse Prediction technique uses many more features of memory accesses to predict the reuse of a given memory access. The features and their parameters are chosen by a stochastic search yielding a

very accurate predictor. For both techniques, the predictions are used to drive a placement, replacement, and bypass optimization that out-performs the state of the art.

### **Short bio**

Elvira Teran is an Assistant Professor in the School of Engineering at Texas A&M International University. Elvira received her doctorate in Computer Science from Texas A&M University in 2017. She is interested in characterizing and exploiting the predictability of programs to improve performance in the memory hierarchy.