Haoyu Chen

Cell: (510) 502-6650 • Email: williamchenthu@gmail.com • Address: 1636 MLK Jr Way, Apt K, Berkeley, CA, 94709

Education

University of California, Berkeley — Overall **GPA: 3.8**
M.S. in **Computer Science** — *December 2014 - December 2015 (expected)*
M.A. in **Statistics** — *August 2013 - May 2014*
Ph.D in Transportation — *August 2011 - December 2015 (expected)*

Tsinghua University — *August 2007 - June 2011*
B.S. in Civil Engineering & Economics (concurrently) — Overall GPA: 92/100; **rank: 1/120**

Relevant Coursework: Natural Language Processing, Machine Learning, Parallel Computing, Data Structure, Algorithm, Machine Structure, Image Manipulation, Time Series, Linear Model, Mathmatical Programming, Behavior Modeling.

Skills

- Proficiency in software development in **Java** and **C**; Familiarity with **Python**, **C++ (OpenMP & OpenCL)**.
- Familiarity with Machine Learning Algorithm (e.g. regression, SVM, and deep neural networks).
- Familiarity with Optimization Algorithm (e.g. stochastic gradient descent, and quasi-Newton method)
- Familiarity with Matlab, R, HTML, LaTeX.

Experience

Graduate student researcher at UC Berkeley (September, 2011 - now)

Accelerate GPU Convolutional Neural Network (deep learning) with Auto-tuning

- Concatenate multiple input images into one batch. Create larger matrix per cuBlas call.
- Create an extra-parameter v, which represents the new size of images. Auto-tune v for peak performance in different layers.
- We speedup the Caffe (Berkeley CNN framework, the fastest open source one as we know) 1.5-2 times.

Accelerate Gibbs parameter estimation on clusters

- Apply butterfly mixing strategy to pass the messages in the cluster network.
- Plan to implement this framework to speedup Gibbs parameter estimation.

Projects and Coursework

Speech Recognition

- Implement speech recognition system based on acoustic model, language model and pronunciation dictionary by Java.
- Construct prefix trie of sub-phones from pronunciation dictionary.
- Use beam model to search sub-optimal state path in the HMM state space and generate recognition results.
- Achieve word error rate (WER) below 35% in 30 mins (RAM $<$ 1.5G) for 200-sentence test.

Probabilistic Parsing

- Build an array-based CKY parser and implement coarse-to-fine pruning by Java. CKY parser achieves a score of 86 F_1 on length 40 sentences. Coarse-to-fine pruning improves the speed by factor of 2.
- Extract the useful features based on parsing tree. Train the classification model by Perceptron, Maximum Entropy or SVM. Rerank the k-best parsing results by classification model, which increases the score to 86 F_1.

Highway Network Congestion Analysis

- Construct probabilistic graphical models (directed graphs) for highway network by Java.
- Compute conditional and marginal vehicle speed probabilities for each node from highway detector data by eliminate (sum-product) algorithm.

Word Alignment

- implement heuristic aligner, IBM Model 1, and HMM model for word alignment between English and French. The HMM model achieves AER 19% with 10000 training sentences.
- Apply the word alignment models in the statistical machine translation system. The translation system achieves BLEU 18 with 10000 training sentences.

Brain Responses to Visual Images

- Predict the response of brain's 3D electromagnetic responses to new photos using machine learning technique by Python.
- Transform the images to feature vectors by Gabor wavelet pyramid.
- Build different regression models to predict the response (ridge, PCR, partial least squares, LASSO). Compare and select the proper model by selection criteria (CV, ES-CV, AIC, BIC).

Stitching Photo Mosaics

- Detect Harris interest points for images, choose proper Harris corners by adaptive non-maximal suppression, and match the detected features.
- Compute the homography transformation matrix with the matched features, and warp the images to the same coordination.
- Blend homographies by gradient-domain fusion and multi-resolution blending to generate photo Mosaics.

LIFC Compiler

- Implement the lexer, parser and code generator for the compiler of homebred LIFC language by C.

Awards

Dwight David Eisenhower Transportation Fellowship, 2013