

Malcolm Greaves

MACHINE LEARNING SCIENTIST AND ENGINEER

San Francisco, CA

☎ (415) 704-9696 | ✉ greaves.malcolm@gmail.com | 🌐 www.malcolmgreaves.io | 📱 malcolmgreaves | 📠 malcolm-greaves-49959919

Experience

Volley Labs, Inc.

San Francisco, CA

NATURAL LANGUAGE UNDERSTANDING SCIENTIST

May 2017 - Present

- Implemented graph-LSTM deep learning model with PSL collective reasoning for automatic knowledge graph creation.
- Research, development, and production deploy of automatic cloze-style, multiple-choice question generation system using a mixture of supervised, unsupervised learning and manually identified lexico-syntactic patterns.
- Technical lead of production AI team. Data pipelines in Airflow; deployed ML models using Keras, Spacy, Tensorflow; Python 3.

Nitro Software, Inc.

San Francisco, CA

RESEARCH ENGINEER

Mar. 2015 - Oct. 2016

- Created novel machine learning based solution for automatic form field detection and semantic classification.
- Lead research, development, and production deployment on first machine learning based product at Nitro.
- Implemented first end-to-end automatic model training and deployment system for Nitro Cloud.
- Full stack production development and deployment in Scala and Javascript with Play!, Angular.js, Postgres, and Kafka.

Alpine Data Labs

San Francisco, CA

SOFTWARE AND MACHINE LEARNING ENGINEER

Jun. 2014 - Mar. 2015

- Implemented machine learning and feature transformation algorithms in Scala as a part of Alpine's analytics and algorithms platform.
- Algorithm optimization for distributed execution on customer's Hadoop and Spark clusters.

Read the Web, Worldly Knowledge Research Groups

Pittsburgh, PA

GRADUATE AND UNDERGRADUATE STUDENT RESEARCHER

Jan. 2010 - May 2014

- Relationship and entity extraction from unstructured text. Includes experience with non-linear learning algorithms, clustering, bayesian inference, and probabalistic graphical modeling.
- Computational linguistics: syntax parsing, part of speech tagging, named entity recognition, sentence segmentation.
- Large scale text classification, feature engineering, vector space modeling, and experience with processing TBs of text on a 175 node Hadoop cluster.

Education

Carnegie Mellon University

Pittsburgh, PA USA

B.Sc. AND M.Sc. IN COMPUTER SCIENCE

Aug. 2009 - May 2014

- Graduated with School of Computer Science Honors
- Four years of research experience in large scale machine learning, natural language processing, and information extraction
- Master's Thesis in semantic relation extraction from unstructured text: <http://goo.gl/DzMr6c>

Work Portfolio

Skills, Knowledge, Programming Languages, Frameworks

- Programming Languages: Proficient in Scala, Python, Go, Java. Moderate fluency in Matlab, R, C
- Production experience: Play! Web Framework, Angular.js, Postgres, Pusher, Apache Spark, Hadoop, Kafka
- Software: Pandas, NumPy, SciPy, TensorFlow, Caffe, BLAS/LAPACK, ScalaNLP Breeze, Spire, Scalaz, Spark (RDD, Pipeline, ML)
- Skills, knowledge: linear algebra, optimization, machine learning, natural language processing and understanding, computer vision and image processing, algorithm development, functional programming, distributed systems, technical communication

smo-fun: Efficient SMO implementation for non-linear SVMs in Scala

[HTTPS://GITHUB.COM/MALCOLMGREAVES/SMO-FUN/](https://github.com/malcolmgreaves/smo-fun/)

- Full implementation of the sequential minimal optimization algorithm. Trains linear and non-linear support vector machines.

fp4ml: Functional programming for machine learning

[HTTPS://GITHUB.COM/MALCOLMGREAVES/FP4ML/](https://github.com/malcolmgreaves/fp4ml/)

- A machine learning library in Scala with clean, functional APIs, strategic, referentially transparent use of mutability, and integration with a dataset type class for painless execution within Spark.