Baojia(Tony) Tong

Cambridge, MA | +1 (617) 710-9767 | tongbaojia@gmail.com
github.com/tongbaojia | tongbaojia.github.io | linkedin.com/in/baojiatonytong

Skills

Languages Python, C++, SQL
Packages PyTorch, DGL, XGBoost, Sklearn, ROOT
Tools Git, Bash, LaTeX, SVN

Experience

Kensho Technologies Cambridge, MA
Tech Lead Apr 2022–present
Machine Learning Engineer Oct 2018–Mar 2022

- Lead the document intelligence team's work and research planning, collaborating with 4 machine learning engineers on OCR, text ordering, and document layout analysis.
- Created and productionized**a novel graph-based deep learning model** for document layout analysis. Achieved state-of-the-art performance in classification and segmentation of documents with less than 10% of the computation cost.
- Combined the graph model with **natural language understanding** to create a key-value extraction tool.
- Invented a novel document layout analysis algorithm and architected it into **a production-level Python package** to **extract tabular and textual** data from PDF documents. The package has been used to extract text and tables from more than **20 million** PDF documents in production.
- Productionized **a financial key-value extraction workflow** for broker research documents. The pipeline automatically extracts $\sim 60\%$ of data with $\sim 98\%$ precision.
- Built and deployed **a general-purpose synonym model** based on Wikipedia open data. The model powers the **text expansion** capability on search platforms daily.

Harvard University Cambridge, MA; Geneva, Switzerland
Ph.D. Student Sep 2012–May 2018

- Searched for double Higgs Boson production at CERN's Large Hadron Collider, introduced novel signal regions to **triple the search sensitivity**, corrected translational modeling effect and **improved background modeling**; published as thesis.
- Implemented a second order correction in Hough Transform extrapolation in C++, **reduced the fake local reconstruction rate** by 50%, and **saved hundreds of hours of computation time** every day.
- **Designed live monitoring software** for reconstruction algorithms and detector performance, inspected and resolved bugs and detector malfunctions within days to maintain data quality.
- Taught undergraduate analytical physics sections, introduction to electronics and experiment analysis in Python, **received two teaching awards** based on student reviews.

Education

Harvard University, Ph.D. in Physics Sep 2012–May 2018
California Institute of Technology, B.A. with honors in Physics Sep 2008–Jun 2012