Wei-Ying Wang, Ph.D.

wayinone@gmail.com   wayinone.github.io   +1-401-556-9066

Melrose, MA (work remotely)

SUMMARY

- Seeking a Senior Data Scientist / Machine Learning Engineer position in which I would contribute to the success of a business

- Data Science Tech Lead in Wayfair, specialized in product matching and classification.

- 5 years of experience of scalable solutions and ML pipeline production

- Applied Mathematics Ph.D.

TECHNICAL SKILLS

Machine Learning	Classification, Regression, Deep Learning, NLP, Approximate Nearest Neighbors
Programming	Python(pyspark, numpy, keras, sklearn, pandas, airflow, tensorflow, kedro)
Database	MSSQL, Vertica, Hive, Big Query

PROFESSIONAL EXPERIENCES AT WAYFAIR LLC.

Product Match

- Matching 60M Wayfair's products against 1.1B crawled products in the market
- Discovering a business rule by analyzing previous matched data, which finds 85% of new matches (8M of match pairs) in 30 min, where previous method will take 3 years
- Developing a 2-staged machine learning model for white-labeled products with 80% AUCPR with Spark ML and Scala
- Constructing an highly scalable ML matching pipeline that can finish entire product matching process in 6 days, while the previous pipeline took 2 months to finish
- Automated retraining and selective deploying to prevent model degradation
- Increasing existing matches by 300% and enabling business analysis of market share, selection gap, and pricing

Product Classification

- Classifying entire catalog (60M Wayfair products and 1.1B crawled competitor products) into 800+ Wayfair classes within 2 hours
- Achieving 90% precision with a language-agnostic deep learning model
- Automated retraining and selective deploying to prevent model degradation

Manufacturer Normalization

- Normalizing different manufacturer synonyms, e.g. "HP", "Hewlett-Packard", and "Hewlett Packard", are alias of the same manufacturer and should be normalized together
- Normalizing 129K distinctive manufacturer names into 10K manufacturers, which covers 97% of crawled data
- Clean and accurate result impacting many aspects company-wide, fueling analysis like product gaps, MSRP estimation, and product matching projects

Part Number Extraction

- Extracting part numbers from 130M competitor's product name and description
- Utilizing a conditional random field model to achieve 95% precision

Optimal Threshold Determination with Bayesian Methods

· Estimating match pair suggestions accuracy by the feedback look from human validation, and utilize Bayesian statistics to obtain a robust result s
· Automatically choosing optimal threshold for different classes of products, and improve overall accuracy by 3%, corresponding to 30K+ hours of labor saving

Code Standardization

· Modifying an open-source template (kedro) to standardize team's code to ensure production-grade coding from the beginning, alleviating data scientists the burden of common infrastrcuture setup like Jupyter Notebook, Spark environment, and code testing
· The first team in Wayfair to advocate the benefit of the code standardization effort with end to end execution
· Speed up develop to production velocity by 200% when launching a new ML pipeline, compared to similar projects carried by the previous team
· Enabling task separation between engineer and data scientist, achieving the maximum efficiency of the team

Image Type Prediction

· Predicting image type (e.g. silhouette, environmental, non-photo, etc.) from product images
· Utilizing Spark with a Tensorflow model to apply prediction on the entire image catalogs (200M images) in 2 hours, as well as predicting new images on daily basis with Airflow
· Saving $450K annual Opex spent on manually tagging images

EDUCATION

Brown University, Providence, RI　　　　　　　　　　　　　　　　　　Sep 2010 - May 2017
Ph.D. Applied Mathematics (GPA: 3.9/4.0)

· Dissertation: Image Compression and Data Clustering: New Takes on Some Old Problems
· Advisor: Stuart Geman

National Taiwan University, Taiwan　　　　　　　　　　　　　　　　　Sep 2004 - May 2006
M.Sc. Mathematics/Track of Statistics (GPA: 3.8/4.0)

National Taiwan University, Taiwan　　　　　　　　　　　　　　　　　Sep 2000 - May 2004
B.A. Economics (GPA: 3.8/4.0)

EMPLOYMENT

Wayfair, Massachusetts　　　　　　　　　　　　　　　　　　　　　　Sep 2017 - Now
Lead Data Scientist / Data Science Manager / Data Scientist

· Developing scalable ML pipeline and developing platform for product image extraction team
· Leading a team of 2 junior data scientists and 1 engineer, creating production pipelines related to product matching

Brown University, Providence, RI　　　　　　　　　　　　　　　　　Jun 2017 - Sep 2017
Postdoc

· Researching the theory behind lossless compression

Academia Sinica, Institute of Mathematics, Taiwan　　　　　　　　Nov 2008 - Aug 2010
Research Assistant

· Researching on an image denoising algorithm

Military Service, Taiwan　　　　　　　　　　　　　　　　　　　　　Jan 2007 - Jan 2008
Coastal Patrol Corporal