Rishabh Misra

rishabhmisra1994@gmail.com +1 (650)-686-9605 linkedin.com/in/misrarishabh rishabhmisra.github.io

EDUCATION

University of California San Diego – Master's in Computer Science **GPA**: 3.93/4.00 **[Sep 2016 – Jun 2018]**

Thapar University, India – Bachelor's in Computer Engineering **GPA**: 9.88/10.00 **[Jul 2011 – Jul 2015]**

RESEARCH EXPERIENCE

➢ **Student Researcher with Prof. Julian McAuley @ UCSD** **[Apr 2017 – Jun 2018]**

- **Decomposing Fit Semantics for Product Size Recommendation in Metric Spaces (Published at RecSys 2018)**
  - Proposed a framework based on latent factor model and metric learning technique to predict fit of different catalog sizes of clothing products so as to recommend appropriately sized products to customers.
  - Contributed the only publicly available datasets for the catalog size recommendation problem.
  - Observed an improvement of up to **18%** over an algorithm developed by Amazon.
- **Fine-Grained Spoiler Detection from Large-Scale Review Corpora (Submitted to ACL 2019)**
  - We contribute a large-scale book review dataset that includes fine-grained spoiler annotations at the sentence-level.
  - Incorporating the findings from exploratory analysis, we develop an end-to-end neural network architecture to detect spoiler sentences in review corpora.
  - Quantitative/qualitative results demonstrate that the proposed method substantially outperforms existing baselines.
- **Modeling Customers' Price-Sensitivity on E-commerce Platform**
  - Initial analysis on Amazon dataset has revealed interesting insights regarding how the ratings of costlier and cheaper alternatives are perceived and how this perception influences customers' purchase decisions.
  - Working towards developing a model that provides price-sensitivity aware recommendations.

➢ **Research Intern @ Indian Institute of Technology, Madras** **[May – Jul 2014, Dec 2014 – May 2015]**

- **Scalable Bayesian Matrix Factorization algorithm (Published at MUSE 2015)**
  - Proposed an MCMC Gibbs sampling algorithm for Matrix Factorization that has linear time complexity with respect to the target rank and linear space complexity with respect to the number of non-zero observations
  - We show empirically that the proposed algorithm performs comparably to Bayesian Matrix Factorization algorithm but runs many orders faster.
- **Scalable Variational Bayesian Factorization Machines (Link to pre-print)**
  - Factorization Machines have Bayesian MCMC sampling-based inference which do not scale on large datasets due to convergence issues.
  - Proposed a scalable variational Bayesian inference-based alternative that performs comparably to the state-of-the-art.

INDUSTRY EXPERIENCE

➢ **Software Development Engineer @ Amazon.com, Seattle, Washington** **[Jul 2018 – Present]**

- Working for Amazon Global team which enables customers to buy products internationally based on export eligibility.
- Current responsibilities include scaling infrastructure and improving eligibility prediction of products.
- **Technologies**: AWS Services | Java | Ruby | Python

➢ **SDE Intern @ Amazon.com, Seattle, Washington** **[Jun 2017 – Sep 2017]**

- Interned at Financial Intelligence Systems team which provides a platform for running big data operational workloads consistently within service level agreement. Worked towards designing and implementing:
  - Support for primary key constraint, batch insert/update and ACID properties while ensuring consistent reads in Hive.
  - Support for non-blocking compaction (carefully discarding old data) to keep read operations efficient.
  - Developed a solution using append-only table and multi-version concurrency control concepts.
- **Technologies**: Java | Hive | DynamoDB

➢ **Member Technical @ Arcesium India Pvt. Ltd., India** **[Jul 2015 – Jul 2016]**

- Worked for the Straight Through Processing team. I was responsible for:
  - Adding support for self-sanitization, self-recovery and fault tolerance in the new infrastructure.
  - Adding self-aware trigger mechanism for EOD trade files, minimizing data completeness issues by 30%.
- **Technologies**: Java | Spring Framework | MyBatis | Microsoft SQL Server

TEACHING EXPERIENCE

➢ **Teaching Assistant @ AWS's Machine Learning University**

- **Introduction to Data Science**
  - **Session**: 1-2019 | **Instructor**: Zachary Levin (Senior Data Scientist)

- **Text Mining**
  - **Session**: 2-2019 | **Instructor**: Pascual Martinez-Gomez (Applied Scientist II)

➢ **Teaching Assistant @ UC San Diego**
- **Recommender Systems and Web Mining (CSE 258)**
  - **Session**: Fall 2017 | **Professor**: Dr. Julian McAuley
- **Software Engineering (CSE 110)**
  - **Sessions**: Fall 2016, Spring 2017, Winter 2018 | **Professor**: Dr. Gregory Kesden, Dr. William Griswold

PROJECTS

➢ **Sarcasm Detection using Hybrid Neural Network** **Python | PyTorch**
- Collected a news headlines-based dataset which improves upon frequently used Twitter datasets by removing the noise in label and language.
- Developed interpretable hybrid neural network architecture with attention mechanism which improves baseline by **5%**.
- Attention module provides insights about the cues that make sentences sarcastic.

➢ **Jointly Modeling Aspects, Ratings and Sentiments with Temporal Dynamics** **Python**
- Implemented a probabilistic graphical framework which utilizes data from product reviews to jointly model aspects of the products, user sentiment on products and associated ratings to predict the unknown ratings.
- For interpretability, model produces insights on the various aspects of products and user sentiment on them which explains the rating.
- Incorporated temporal information into the joint model which improves performance by **1%** and additionally provides insights into how users' preference of different product aspects change over time.

➢ **Hierarchical Attention Network for Rating Prediction** **Python | Keras**
- Implemented a hierarchal RNN with attention mechanism that uses product reviews to predict the product ratings.
- Attention mechanism allows the RNN to focus on words and sentences that best explain the rating given to an item and uses this knowledge to predict unknown ratings.

➢ **Ensemble of CNNs for Traffic Lights Recognition** **Python | Keras**
- Developed an ensemble of CNNs where each CNN is built with the motive of giving a superior performance while keeping the model size small. Small model size allows for quick training even when computational resources are scarce.
- We train the model on Nexar traffic lights challenge dataset and show that it achieves >90% detection accuracy.
- We also performed localization experiments using Faster R-CNN on UCSD traffic lights dataset to demonstrate high performance of the model for object detection tasks.

➢ **Music Generation using Character-level RNN** **Python | Keras**
- Proposed a character-level RNN that learns the structure of music files in ABC format and generates music from the trained network in the ABC format, which can then be converted into audio format.
- We experiment with different parameters like temperature, dropout, the number of neurons in the hidden layer and the effect of optimizers and observe their effect on model's learning behavior.

➢ **Hotel Recommendation System** **Python | Pandas | Scikit-learn**
- This recommender system aims to recommend top five most likely hotels that customers might want to stay in.
- The proposed model is an ensemble of Random Forest, Naïve Bayes, SGD classifier and XGBoost and utilizes the Expedia Hotel Recommendation dataset for training and evaluation.

➢ **Review Helpfulness Prediction** **Python | Scikit-learn**
- Performed feature engineering on Amazon dataset to identify best features that best correlate to helpfulness of reviews.
- Proposed a linear regression-based model with polynomial interaction between features that ranked in top **12%** of the solutions on Kaggle.

ACHIEVEMENTS AND POSITIONS OF RESPONSIBILITY

➢ Received a **financial grant** from Python Software Foundation to attend PyCon 2019 in Cleveland, Ohio.
➢ **Program committee member** for the SciPy 2019 conference.
➢ **Reviewer** for Amazon's Machine Learning Conference (**AMLC**) 2019.
➢ **Technical writer** for the Towards Data Science online publication.
➢ **Mentored** first-generation undergraduate students at UCSD under the Jacobs Undergraduate Mentorship Program.
➢ **Mentored** newly joined graduate students at UCSD as part of Graduate Women in Computing community.
➢ Received **university medal** for being the topper of Computer Science batch at Thapar University
➢ Received **full merit scholarships** for undergraduate studies at Thapar University.
➢ Received **certificate of merit** for Mathematics from CBSE for being in top 0.1% of students in 12th standard.
➢ Qualified for **ACM-ICPC** Asia Region, Kanpur site 2013 held at IIT Kanpur.