Allen (Shuo) Lin
al001@tamu.edu

Education

Texas A&M University College Station, TX
Doctor of Philosophy in Computer Science July 2021 - present
Advisor: James Caverlee

Duke University Durham, NC
Master of Science in Computer Science Aug. 2019 – May 2021
Advisor: Allison J.B. Chaney
Thesis: Disentangled Representation Learning for Neural Collaborative Filtering

Ohio State University Columbus, OH
Bachelor of Science in Computer Science Aug. 2015 – May 2019
National Buckeye Scholarship (half tuition full terms)

Research Experiences

Additive Poisson Factorization *May. 2020 – Aug. 2020*

- Designed and implemented a variational inference framework, specifically for Poisson factorization based conjugate models, that derives update procedures and performs coordinate ascent optimization for all latent variables, given their priors and generative processes
- Validated our variational inference framework by showing it outperforms the state-of-the-art variational inference libraries, Stan and Edward, in both speed and recommendation accuracies measured in NDCG

Disentangled Representation Learning for Neural Collaborative Filtering *Sep. 2020 – Mar. 2021*

- Modified the β -VAE disentanglement metric initially proposed by Google to make it applicable to neural collaborative filtering.
- Proposed a complementary disentanglement metric that exams each latent factor in learned user and item representations shares no mutual information with any other.
- Implemented an end-to-end pipeline for learning disentangled user and item representation in the context of neural collaborative filtering.
- Assessed the relationship between β and recommendation performance; the relationship between β and the degrees of disentanglement achieved in learned user and item representations; and the relationship between latent dimensionality and recommendation performance.

Semantic Image Segmentation - *Duke Center for AI in Radiology* *Nov. 2020 – Mar.2021*

- Conduct a comparative analysis to determine the best type of week annotations (e.g., image-level labels, point, scribble, bound box, polygon) for semantic segmentation under a fixed budget constraint
- Design pixel-wise mask deformation algorithms that simulate both intentional and unintentional errors commonly made by human annotators then examine the robustness of current benchmarking segmentation architectures (e.g., DeepLab v3+, U-Net, SegNet, FCN) to such noisy masks

Relevant Courses

- Probabilistic Machine Learning
- Deep Learning
- Applied Natural Language Processing
- Data Mining
- Advanced Artificial Intelligence
- Matrices and Vector Spaces
- Machine Learning
- Intelligent UI

Relevant Projects

Deep Fake Face Detection (*Python, Pytorch*) *Sep. 2020 – Nov.2020*

- **Binary Representation Learning**: Implemented a β variational autoencoder that respectively learns the latent representations for both the real and the fake face images.
- **Binary Representation Classification**: Trained a random forest classifier on the dataset composed of all learned latent representations where all representations of real face images are labeled as 1 and all representations of fake face images are labeled as 0.
- **One-Class Anomaly Detection**: Trained a variational autoencoder using only the real face images, then threshold the reconstruction error to determine if an unseen face image is fake or real

Multilingual Hierarchical Patent Classification (*Python*) *Feb.2020 – Apr. 2020*

- **Transfer Learning**: Incorporated and tuned a pre-trained BERT multilingual model consisted of 110M parameters to achieve an F1 score of 0.5 where the baseline model achieved an F1 score of less than 0.1
- **HARNN**: Implemented a hierarchical attention recurrent model that combines attention-based mechanism and Bi-LSTM to capture semantic information

IEEE Credit Card Fraud Detection Competition (*Python, PyTorch*) *Sep. 2019 – Dec. 2019*

- **Class Imbalance Handling**: Performed under-sampling on the non-fraudulent samples using Tomek-Links and Cluster Centroids and oversampling on the fraudulent samples using SMOTE and Generative Adversarial Network to counter the severe class imbalance problem effectively
- **Dimensionality Reduction**: Implemented the elbow method to determine the optimal number of features to keep and then performed PCA to reduce the dimensions of the augmented dataset
- **Machine Learning Models**: Implemented a Stacked Autoencoder, a Bidirectional Multilayer Recurrent Neural Network, a Stacked Restricted Boltzmann Machine, and a Random Forest classifier
- **Evaluation**: Achieved an accuracy of 97.4% when classifying unseen real-world credit card transactions

Publications

- **CIKM 22** **Allen Lin**, Jianling Wang, Ziwei Zhu, and James Caverlee. Quantifying and Mitigating Popularity Bias in Conversational Recommender Systems. The 31st ACM International Conference on Information and Knowledge Management, 2022.
- **RecSys 22** **Allen Lin,** Ziwei Zhu, Jianling Wang, and James Caverlee. Towards Fair Conversational Recommender Systems. The 16th ACM Conference on Recommender Systems, 2020. (workshop paper)