Yinkai Wang

18650146958 | ywang88@gmu.edu
yinkaiw.github.io

EDUCATION

George Mason University Aug 2018 - Aug 2021
Computer Science Bachelor VSE bachelor Fairfax
- Honors/Awards: Dean's List(2018-2020)

Huaqiao University Aug 2017 - Aug 2022
Computer Science Bachelor VSE bachelor Xiamen

PUBLICATIONS

- **Yinkai Wang***, Kaiyi Guan*, Aowei Ding*, Yuanqi Du. Ensemble Machine Learning System for Student Academic Performance Prediction. Educational Data Mining (EDM) 2021, Workshop for Undergraduates (W4U).
- Yuanqi Du, **Yinkai Wang**, Fardina Alam, Yuanjie Lu, Xiaojie Guo, Liang Zhao, Amarda Shehu. Deep Latent-Variable Models for Controllable Molecule Generation. Research Paper for IEEE BIBM 2021. (In submission)
- **Yinkai Wang***, Kaiyi Guan*, Aowei Ding*, Shixi Wu*, Yuanqi Du.Improved Student Performance Prediction via Ensemble Machine Learning. Workshop for Neurips 2021(In submission)
- **Yinkai Wang**, Antonis Anastasopoulos. On the Cross-Lingual Consistency of Named Entity Recognition Models. Student Abstract for AAAI-UC. (In submission)
- Fahim Faisal, **Yinkai Wang**, Antonis Anastasopoulos. Dataset Geography: Mapping Language Data to Language Users. Research paper for ARR2021. (In submission)

RESEARCH EXPERIENCE

Multilingual Geospatial Language Expression Discovery
Research Assistant(Advisor: Prof. Antonios Anastasopoulos)
- Take parallel data between English and another language, Use existing tools to discover GLEs in English.
- Align the parallel English--X data, Project the labels from English to X.
- Compare the projected annotations on cmn,ell,ita with the ones produced by the in-language models.

Diffusion Probabilistic Models for Protain Generation
Research Assistant(Advisor: Dr. Amarda Shehu)
- Use diffusion probabilitic models to generated protain.
- Use short range and lang range to evaluate the result.

Predicting Minimum Inhibitory Concentration for Quaternary Ammonium Compounds w/ Machineing Learning
Research Assistant(Advisor: Dr. Amarda Shehu)
- Create three seeting based on the 70 features of ~450 Quaternary Ammonium Compounds.
- Predict four Minimum Inhibitory Concentration value with ten regression machine learning models.

Ensemble Machine Learning System for Student Academic Performance Prediction
Researcher
- Used ensemble machine learning system to predict students' final grades.
- Due to the influence of COVID-19, the education to students faced a severe problem: It's much harder for teachers to help students and know students' learning conditions.
- This model consists of two components, the ensemble feature engineering module, and the ensemble prediction module. Extensive experiment results have shown the superiority of our model over other traditional machine learning models, both in stability, efficiency, and accuracy.

Deep Latent-Variable Models for Controllable Molecule Generation
Research Assistant(Advisor: Dr. Amarda Shehu)
- Proposed several deep latent-variable models to generate small molecules with desired molecular properties.
- The models operate under supervised, disentangled representation learning and leverage both graph representation learning to learn inherent constraints in the chemical space and inductive bias to connect chemical and biological space.
- The evaluations show that the models are a promising step in controllable molecule generation in support of cheminformatics, drug discovery, and other application settings.

PROFESSIONAL EXPERIENCE

Bytedance Apr 2021 - Jul 2021
Intern DevEco Beijing
- Focus on the base of hostapp of android, which have the coupling relationshiop with most of app from bytedance, like tiktok, lark, Toutiao...
- Create a mock setting enviroment implement to help QA test, greatly improve the efficiency of testing and publishing.
- Handle unexpected technology problems of user. Judge the urgency of problems and solve it.
- Make a great and comfortable enviroment for all the deleloper who are developing microapp on bytedance.

Google Kaggle
Google Smartphone Decimeter Challenge
- Ranked top 20% on the Google Smartphone Decimeter Challenge w/ a group of five undergraduate students.
- Designed data clearning, preprocessing, data analysis, model selection, result evaluation and visualization pipeline.
- Mastered real-world data science challenge with machine learning pipeline and team collaboration.

SKILLS LIST

Programming skills: Python, Java, C, Kotlin, MIPS, Julia
Research Interests: Machine Learning, Deep Learning, AI for Science, Deep Graph Learning, Natural Language Processing