Changwen Xu

412-999-1188 | changwex@andrew.cmu.edu | linkedin.com/in/changwenxu | github.com/ChangwenXu98

EDUCATION

Carnegie Mellon University   Pittsburgh, PA
Master of Science in Computational Materials Science and Engineering, GPA: 4.0/4.0   *Dec. 2022 (Expected)*

South China University of Technology   Guangzhou, China
Bachelor of Engineering in Materials Science and Engineering, GPA: 3.96/4.0   *Jun. 2021*

PUBLICATION

1. **Xu, C.**, Wang, Y., Barati Farimani, A., TransPolymer: a Transformer-based Language Model for Polymer Property Predictions. *arXiv preprint arXiv:2209.01307*, 2022 (reviewed by *npj Computational Materials*)

2. Wang, L. (#), **Xu, C.** (#), Zhang, W., Zhang, Q., Zhao, M., Zeng, C., Jiang, Q., Gu, C., Ma, Y., Electrocleavage Synthesis of Solution-Processed, Imine-Linked, Crystalline Covalent Organic Framework Thin Films, *J. Am. Chem. Soc.*, 2022, 144, 20, 8961–8968. (# These authors contributed equally.)

RESEARCH EXPERIENCE

Graduate Student Researcher   Pittsburgh, PA
Carnegie Mellon University   *Oct. 2021 – May. 2022, Aug. 2022 – Present*

Graduate Research Assistant   Pittsburgh, PA
Carnegie Mellon University   *May. 2022 – Aug. 2022*

Score-based Generative Model for Time-integrated Molecular Dynamics | *Python, PyTorch, MD*
- Build **score-based** generative model for predicting MD trajectories of water molecules by conditional generation
- Perturb water molecules with **conditional noise** and employs previous MD snapshots as the **prior distribution** for sampling
- Integrate **equivariant Transformer** to approximate the diffusion process and train the model to predict longer trajectories in an **autoregressive** manner

Transformer-based Language Model for Polymer Property Predictions | *Python, PyTorch, Language Model*
- Built **TransPolymer**, a **Transformer** based model for accurate and efficient polymer property prediction
- Conducted literature review on molecular language models, collected datasets from literature, and designed a tokenization strategy with chemical awareness
- **Pretrained** with a **masked language model** on a large unlabeled dataset of polymer sequences for learning expressive representations
- **Finetuned** the pretrained model in different downstream prediction tasks, achieving the state-of-the-art (**SOTA**) results and surpassing baseline models by large margins in most cases
- Demonstrated the effectiveness of pretraining by **t-SNE**, and identified influential parts of polymers on prediction results by **attention scores visualization**
- Highlighted the model as a promising computational tool for understanding chemistry in a data science view
- Wrote an article about this project which is under review by *npj Computational Materials*

Undergraduate Student Researcher   Guangzhou, China
South China University of Technology   *Sep. 2018 – Jun. 2021*

Electrocleavage Synthesis of Solution-Processed Covalent Organic Framework (COF) Thin Films
- Developed an unprecedented electrocleavage synthesis strategy to produce **imine-linked COF films** directly on electrodes from electrolyte solutions at room temperature
- Carried out **cathodic exfoliation** of COF powders into nanosheets by electrochemical reduction and protonation and **anodic oxidation** to reproduce COF structures
- Characterized the COF films, COF solutions as well as precipitated COF nanosheets to verify the method
- Published a journal article about this project in *Journal of the American Chemical Society* as the co-first author

Synthesis and Solution Processing of Triazine-based Crosslinked Conjugated Polymers (TCPs)
- Reported the method of **protonating** the polymeric skeletons to enhance the solubility of TCPs under electrostatic repulsions to achieve solution processing

- Synthesized triazine-based n-type crosslinked conjugated polymers with **perylene diimide** units and achieved **solution-processing** of crosslinked conjugated polymers for the fabrication of polymeric films
- Developed the method of electrocleavage synthesis of triazine-based COFs from previous research and compared the results of the two methods.
- Completed the graduation thesis on this project and was awarded "**Outstanding Undergraduate Thesis**" of South China University of Technology

PROJECTS

AI Ramanujan: Discovery of Formula Equivalence (CMU) | *PyTorch, Language Model*   Oct. 2022 – Present
- Train **Transformer** encoders to evaluate whether two mathematical formulas map to the same value
- Construct a **tokenizer trainer** for adaptation to formulas in Mathematica format and design a **translator** to convert formulas in Maple format into Mathematica format
- Encode formulas as texts and **Abstract Syntax Trees** to learn both the lexical and syntactic information from formulas
- Adapt the encoders with **cross attention** to encode correlations between two embeddings and train the model by **contrastive learning**

JPX Tokyo Stock Exchange Prediction (CMU) | *Python, PyTorch, sklearn, EDA*   May. 2022 – Jul. 2022
- Trained **LSTM** model for JPX stock prediction and compared the model against real future returns for evaluation
- Applied exploratory data analysis on JPX stock data, conducted data cleaning, and carried out **feature engineering** to construct financial descriptors
- Achieved the highest rank of top 15% in Kaggle competition "JPX Tokyo Stock Exchange Prediction"

Developing disaster relief and response system (SCUT) | *Python, MATLAB*   Jan. 2019 – Jan. 2019
- Participated in 2019 Mathematical Contest in Modeling working on developing a feasible aerial disaster relief and emergency response system, and was designated as Meritorious Winner
- Determined the optimal drone fleet, payload packing configurations, and cargo container locations of the system
- Proposed **elliptic interpolation algorithm** to design delivery routes and schedules of flight plan

INVITED TALKS

Invited Speaker at "Math + Machine Learning + X" Seminar at Brown University   Nov. 2022
- Gave a talk on *TransPolymer: a Transformer-based Language Model for Polymer Property Predictions*

HONORS AND AWARDS

Outstanding Undergraduate of MSE, South China University of Technology	Jun. 2021
Xiaolan Wu Scholarship	Jun. 2021
Outstanding Undergraduate Thesis	Jun. 2021
Kemingda Scholarship	Oct. 2020
Hongping Changqing Foundation Scholarship	Jun. 2019 – Jun. 2020
National Scholarship (Twice) (0.2% among all the university students in China)	Dec. 2018 – Dec. 2019
First Prize (Guangdong) of China Undergraduate Mathematical Contest in Modeling	Nov. 2019
Meritorious Winner of Mathematical Contest in Modeling	Apr. 2019
Second Prize of "FLTRP" Cup English Writing Contest	Dec. 2018

SKILLS

Computational Skills:
  Programming Languages: **Python**, SQL, MATLAB, R
  Frameworks: **PyTorch**, Tensorflow, Keras
  Developer Tools: Google Cloud Platform, PyCharm, VS code, Visual Studio, Docker, Git
  Libraries: pandas, NumPy, Matplotlib, sklearn, Huggingface, RDKit
  Deep Learning Model: **Transformer**, **DDPM**, **Score-based Generative Model**, LSTM, CNN
  Computation and Simulation Tools: **Molecular Dynamics**, DFT
Experimental Skills:
  Synthesis Skills: organic synthesis of small molecules and COFs, purification skills like column chromatography and Soxhlet purification
  Characterization Techniques: FTIR, UV-vis spectrophotometry, XRD, SEM, XPS
  Electrochemical Experimental Skills: CV, DPV, chronoamperometry
Additional Skills:
  Spoken Languages: English (fluent), Chinese (native)