Langlin Huang

Email:h.langlin@wustl.edu | **GitHub**:https://shrango.github.io/

EDUCATION

•**Washington University in St. Louis**   *St.Louis, U.S.*
Ph.D. Student in Computer Science   Aug. 2024 - present

•**Institute of Computing Technology, Chinese Academy of Sciences (ICT/CAS)**   *Beijing, China*
M.E. in Computer Technology -GPA: 3.75/4.0, Outstanding Dissertation   Sep. 2021 - Jun. 2024

•**University of International Business and Economics**   *Beijing, China*
B.E. in Data Science and Big Data Technology -GPA: 3.74/4.0 (Rank:2/147)   Sep. 2017 - Jun. 2021

RESEARCH INTERESTS

- Language Model Reasoning & Efficient Model Generation
- Machine Translation & Multilingual Representation Learning

PUBLICATIONS

•**MoCE: Adaptive Mixture of Contextualization Experts for Byte-based Neural Machine Translation**
Langlin Huang, Mengyu Bu, Yang Feng
NAACL 2025 oral [Paper] [Code]

•**Integrating Multi-scale Contextualized Information for Byte-based Neural Machine Translation**
Langlin Huang, Yang Feng
ACL findings, 2024 [Paper][Code]

•**Enhancing Neural Machine Translation with Semantic Units**
Langlin Huang, Shuhao Gu, Zhuocheng Zhang, Yang Feng
EMNLP findings, 2023 [Paper][Code]

•**BayLing: Bridging Cross-lingual Alignment and Instruction Following through Interactive Translation for Large Language Models**
Shaolei Zhang, Qingkai Fang, Zhuocheng Zhang, Zhengrui Ma, Yan Zhou, **Langlin Huang**, Mengyu Bu, Shangtong Gui, Yunji Chen, Xilin Chen, Yang Feng
Preprint [Paper] [Code]

•**Divide, Reweight, and Conquer: A Logit Arithmetic Approach for In-Context Learning**
Chengsong Huang, **Langlin Huang**, Jiaxin Huang
ICLR 2025 workshop [Paper][Code]

•**Efficient test-time scaling via self-calibration**
Chengsong Huang, **Langlin Huang**, Jixuan Leng, Jiacheng Liu, Jiaxin Huang
Preprint [Paper][Code]

•**CrossWordBench: Evaluating the Reasoning Capabilities of LLMs and LVLMs with Controllable Puzzle Generation**
Jixuan Leng, Chengsong Huang, **Langlin Huang**, Bill Yuchen Lin, William W Cohen, Haohan Wang, Jiaxin Huang
Preprint [Paper][Code] [Dataset]

•**Automatic Construction of a Depression-Domain Lexicon Based on Microblogs: Text Mining Study**
Genghao Li, Bing Li, **Langlin Huang**, Sibing Hou
JMIR medical informatics, 2020, Vol 8. Jun. 2020 [Paper]

PROJECTS

•**CVAE-based Label Smoothing for Neural Machine Translation**   *Feb. 2022 - Aug. 2022*
Proposed a flexible label smoothing for training language models and translation models.

- **Contributions**: Proposed to replace uniform distributions with predicted real label distributions in label-smoothed cross-entropy loss.
  Proposed to predict real label distribution with a Conditional Variational Auto Encoder(CVAE) module by foreseeing the ground truth word.
- **Achievement**: Significantly improved translation performance by 1.2 BLEU on En-Ro and Zh-En translation tasks.
- **Patent Link**: http://epub.cnipa.gov.cn/patent/CN115455993A

•**Chinese-Thai Translation System**   *May. 2022 - Jul. 2022*
Developed strong Chinese-Thai bidirectional machine translation systems.

- **Contributions**: Proposed a strategy to modify pre-trained language model mBART, without hurting performance.
  Crawled external in-domain texts and augmented training data via back-translation.
- **Achievement**: Won the **Championship** in the 18th China Conference on Machine Translation(CCMT) Zh-Th track.
- **Technical Report link**: http://sc.cipsc.org.cn/mt/conference/2022/papers/test_paper/60/60_Paper.pdf