

Junjie Chen

✉ junjiechen.chris@outlook.com | 🌐 https://chrisjic.com/ | 📧 junjiechen-chris | 📄 junjie-chen-b6a075240

个人总结

计算机科学博士，拥有5年NLP经验。拥有丰富的大模型应用研究经验。从零开始打造多模态整书翻译系统及其自动评价系统。基于大模型的无监督句法分析研究被ICLR2025选为spotlight。熟悉NLP，语音处理，CV相关的AI/ML方法知识。

工作经历

Amazon

东京，日本

AMAZON SCIENCE FELLOW

2025年8月 - 2026年2月（预计）

- 多模态整书翻译项目的核心成员与贡献者
- 设计并落地人工评价流程
- 开发基于大模型的自动评测系统，显著提升评测准确率
- 开发以大模型为核心的AI翻译流水线，显著提升翻译质量

乐天株式会社 (Rakuten)

东京，日本

研究实习生

2024年11月 - 2025年2月

- 调研大模型稀疏自编码器 (SAE) 的业务应用场景
- 将基于SAE的个人信息识别准确率从72%提升至87%

国立情报学研究所 (NII)

东京，日本

研究助理

2024年6月 - 2025年2月

- 搭建针对180M至175B不同规模日语大模型的评测流程
- 维护800张GPU规模的训练集群

东京大学

东京，日本

教学助理

2022年10月 - 2023年2月

- 指导3个学生团队完成NLP期末项目（其中一项成果发表于JNLP-2023）

技能

研究方向	基于大模型的无监督句法分析、多模态整书翻译
生成建模	语言模型、扩散模型、正规化流、变分自编码器
机器学习框架	Pytorch、Tensorflow、Lightning、Hydra、vLLM、Bedrock
分布式计算	Slurm、Huggingface Accelerate、Megatron、Lightning
DevOps	AWS、Cloudflare、Docker
编程技能	Python、Typescript、Bash
语言能力	中文（母语）、日语（N1，商务）、英语（C1，商务）

研究项目

Amazon

整书翻译项目

2024 - 2025

- 从零制定基于MQM的翻译评测指南，并开发一系列基于大模型的自动评测系统
- 构建以大模型为核心的翻译系统。以单一系统整合视觉定位、对话框检测、剧情理解与翻译等子功能
- 将标注一致性从0（无一致）提升至0.3（一般一致）
- 将页均翻译MQM（越低越好）从约15降至约3，降低翻译错误率逾5倍
- 将自动评测精确率从13%提升至45%，召回率从5%提升至50%。
- 改进对话框检测效果，显著提高召回同时保留高检测精度。

东京大学，博士阶段研究项目（二）

基于语义信息发现的无监督成分句法分析

2024 - 2025

- 提出基于释义的字符串袋模型，用于估计字符串级别的语义信息
- 设计校准训练目标，适用于基于PCFG的无监督成分句法分析
- 训练目标在4种语言上平均带来8%的绝对精度提升
- 语义信息估计在识别基本语义单元方面有效

东京大学，博士阶段研究项目（一）

结合语法先验的语言模型依存句法分析

2023 - 2024

- 提出基于 Metropolis-Hasting 的词间互信息估计算法
- 提出子采样方法，将语法先验融入互信息计算
- 语法先验带来超过 5% 的整体准确率提升
- 在语义相关依存预测上实现超过 10% 的精度提升

东京大学，硕士阶段研究项目

结合混合专家模型的语义依存分析

2021 - 2022

- 设计混合专家模型，根据句法模式建模不同的语义依存关系
- 应用变分推断自动聚类相似的句法模式
- 相比最新基线在语义依存分析上显著提升性能

教育背景

东京大学，导师：Yusuke Miyao

东京，日本

博士，计算机科学

2021 年 10 月 - 2025 年 3 月

- 获日本学术振兴会 (JSPS) 奖学金资助
- 博士研究成果发表于 3 个顶级会 (NAACL 2024/ACL 2024 Findings/ICLR 2025 (Spotlight))

东京大学，导师：Yusuke Miyao

东京，日本

硕士，计算机科学

2019 年 10 月 - 2021 年 9 月

- 硕士研究成果发表于 ACL 2022 (Oral)

利物浦大学，导师：Danushka Bollegala

利物浦，英国

访问学者

2023 年 10 月 - 2024 年 3 月

- 研究成果发表于 ACL 2024 Findings

东北大学

中国，沈阳

学士，软件工程

2015 年 9 月 - 2019 年 6 月

- GPA: 3.46/4.0 (北海道大学交换 6 个月期间 GPA 3.78/4.0)

精选论文

Improving Unsupervised Constituency Parsing via Maximizing Semantic Information

ICLR 2025 (Spotlight, 前 5.1%)

JUNJIE CHEN, XIANGHENG HE, YUSUKE MIYAO, AND DANUSHKA BOLLEGALA

2025

Constituents are Frequent Word Sequences among Sentences with Equivalent Predicate-Argument Structures: Unsupervised Constituency Parsing by Span Matching

ACL 2024 Findings

JUNJIE CHEN, XIANGHENG HE, DANUSHKA BOLLEGALA, AND YUSUKE MIYAO

2024

Language Model Based Unsupervised Dependency Parsing with Conditional Mutual Information and Grammatical Constraints

NAACL 2024

JUNJIE CHEN, XIANGHENG HE, AND YUSUKE MIYAO

2024

Syntactic-Semantic Dependency Correlation in Semantic Role Labeling: A Shift in Semantic Label Distributions

Journal of Natural Language Processing

JUNJIE CHEN

2022

Modeling Syntactic-Semantic Dependency Correlations in Semantic Role Labeling Using Mixture Models

ACL 2022 (Oral)

JUNJIE CHEN, XIANGHENG HE, AND YUSUKE MIYAO

2022

ProsodyFM: Unsupervised Phrasing and Intonation Control for Intelligible Speech Synthesis

AAAI 2025 (Oral)

XIANGHENG HE, JUNJIE CHEN, ZIXING ZHANG, BJÖRN W SCHULLER

2025

Task Selection and Assignment for Multi-modal Multi-task Dialogue Act Classification with Non-stationary Multi-armed Bandits

XIANGHENG HE, JUNJIE CHEN AND BJORN W. SCHULLER

ICASSP 2024

2024

An Improved StarGAN for Emotional Voice Conversion: Enhancing Voice Quality and Data Augmentation

XIANGHENG HE, JUNJIE CHEN, GEORGIOS RIZOS, AND BJÖRN W. SCHULLER

Interspeech 2021

2021

A system for worldwide COVID-19 information aggregation

AKIKO AIZAWA, FREDERIC BERGERON, JUNJIE CHEN, FEI CHENG, KATSUHIKO HAYASHI, KENTARO INUI, HIROYOSHI ITO, DAISUKE KAWAHARA, MASARU KITSUREGAWA, HIROKAZU KIYOMARU, MASAKI KOBAYASHI, TAKASHI KODAMA, SADAOKUROHASHI, QIANYING LIU, MASAKI MATSUBARA, YUSUKE MIYAO, ATSUYUKI MORISHIMA, YUGO MURAWAKI, KAZUMASA OMURA, HAIYUE SONG, EIICHIRO SUMITA, SHINJI SUZUKI, RIBEKA TANAKA, YU TANAKA, MASASHI TOYODA, NOBUHIRO UEDA, HONAI UEOKA, MASAO UTIYAMA, YING ZHONG

Workshop on NLP for COVID-19

2020

A pattern-based method for medical entity recognition from Chinese diagnostic imaging text

ZIHONG LIANG, JUNJIE CHEN, ZHAOPENG XU, YUYANG CHEN, TIANYONG HAO

Frontiers in Artificial Intelligence

2019

A Bibliometric Analysis of the Research Status of the Technology Enhanced Language Learning

XIELING CHEN, JUNTAO HAO, JUNJIE CHEN, SONGSHOU HUA, TIANYONG HAO

Emerging Technologies for Education

2018

资助与奖学金

- 2023 DC2 奖学金, 日本学术振兴会 (JSPS)
- 2024 杰出学生特别津贴, 日本学术振兴会 (JSPS)
- 2025 参会差旅资助 (\$2000), 美国人工智能协会 (AAAI)
- 2022 IST-RA 奖学金, 东京大学