

Andong Chen

+86 155-2646-6911 ands691119@gmail.com
https://andongblue.github.io/chenandong.github.io/

EDUCATION

Mizu University of China	Sep 2019 - Jun 2022
Natural Language Processing, Master Information Technology	Beijing
Queen Mary University of London(International Exchange&Double Degree)	Jan 2021 - Jan 2022
Big Data Science, Master Science and Engineering	London
Hunan University	Sep 2015 - Jun 2019
Electrical Engineering and Automation, Bachelor College of Electrical and Information Engineering	Chang Sha

PUBLICATIONS&PATENT

JCapsR: A joint capsule neural network learning model for Tibetan language knowledge graph representation	Jan 2021 - May 2021
China and National Language Resource and Monitoring Research Center	Beijing

Introduction: In this paper, we propose a joint capsule neural network (JCapsR) learning model for Tibetan language knowledge graph representation based on the previously constructed Tibetan language knowledge graph.

- Built the TransR and Caps models and completed the whole experiment
- Wrote the experimental section

A Joint Model for Representation Learning of Tibetan Knowledge Graph Based on Encyclopedia	Dec 2019 - Feb 2021
China and National Language Resource and Monitoring Research Center	Beijing

Introduction: The paper mainly constructs the first Tibetan encyclopedic knowledge graph dataset and improves the representation ability of the Tibetan knowledge graph.

- Built the TransE and CNN models and completed the whole experiment
- Wrote the whole paper with Prof. Yuan Sun

Patent : A service system for automatic generation of reading questions in Tibetan language for elementary school	Jun 2021 - Jun 2021
Minzu University of China	Beijing

Introduction: The present invention relates to a service system for the automatic generation of reading questions in the Tibetan language for elementary school.

Patent Number: CN202110228195.5

- Extract the features from the Tibetan texts of each grade by logistic regression as a classification model
- Obtain a large number of Tibetan texts that can be read by primary school students through crawlers and the classification model
- Fine-tune T5 model and obtain the Tibetan question generation machine

OPEN SOURCE CONTRIBUTION

Thorough-pytorch	Jul 2021 - Present
Role: Content builder	Beijing

Introduction: Through team learning, you will be able to master the basic knowledge and content of PyTorch from shallow to deep.

- Result: **454 star** and **302 fork**
- Open source link: <https://github.com/datawhalechina/team-learning-program/>

Food Voice Recognition(Alibaba-Tianchi-Match)	Jan 2021 - May 2021
Role: Competition designer and content builder	HangZhou

Introduction: Alibaba-Tianchi and Datawhale Co-Sponsor Basic Data Science Competition

- In Ali Tianchi. I designed the competition, provided speech recognition knowledge, written Baseline, and live streaming at Tianchi to interpret the competition questions and codes.

Docker-Team-Learning	Jan 2021 - Mar 2021
Role: Content builder Datawhale	Beijing

Introduction: This tutorial is a collaborative project with “docker from beginner to practice”, and we

have reorganized and refactored it based on the original project with the permission of the author.

- Result: **454 Start** and **302 fork**

- Open source link: <https://github.com/datawhalechina/team-learning-program/>

Hands-on data analysis

Jan 2019 - May 2019

Role: Project sponsor and content builder

Beijing

Introduction: An open-source data analysis project for everyone to learn by themselves

- Use the Titanic data set to divide tasks according to data observation, data manipulation, data visualization, data modeling, and model evaluation.

- Result: **400 start** and **163 fork**

- Open source link: <https://github.com/datawhalechina/hands-on-data-analysis>

INTERNSHIP EXPERIENCE

Institute of Computing Technology, Chinese Academy of Sciences

May 2021 - Present

NLP Research Intern Ai Lab

Beijing

Introduction 1: The project mainly does quantification of deep learning models

- Do quantification of the speech model Waveglow and evaluate the quality of quantification by crowdsourcing.

- Build the first NLP quantitative reasoning benchmark with mentors.

- Compare with Nvidia's Transformer quantization solution.

Introduction 2: Build a large pre-trained language model

- Choose a better token vocabulary through different experiments.

- Research position details and chooses a better way.

- Build data cleaning and de-duplication tools.

Queen Mary University of London

Feb 2021 - Present

NLP Research Intern Computational Linguistics Lab

London

Introduction: Research low-resource language machine reading comprehension

- The experimental part has been completed.

- The paper will be submitted to ACL.

HONORS & AWARDS

Third Prize Scholarship (10%), Mizu University of China ,

2019-2020

“Huawei Cup” China Post-Graduate Mathematical Contest in

2020

Modeling(17th) , Third Prize(15%) , The Chinese Society of Academic Degrees and Graduate Education

China College Students’ ‘Internet+’Innovation and Entrepreneurship

2020

Competition(6th), Third Prize(25%), Beijing Municipal Education Commission

WeBank’s fintech university competition(2th), Top 10(10/100),

2020

WeBank & Shenzhen University WeBank Institute of Fintech

TALKS

Alibaba Cloud-TianChi: Speech Recognition Novice Competition

Jun 2021

Invited Talk, Alibaba Cloud-TianCh, Beijing

Open-source: Open source yourself

Aug 2021

Invited Talk, Institute of Computing Technology, Chinese Academy, Beijing

Paper Share: ERNIE 3.0

Aug 2021

Invited Talk, DatawhalePaper, Beijing

Bert and Machine Reading Comprehension

Jul 2021

Invited Talk, Datawhale and Graviti, Beijing

SKILLS

- **Programming Languages:** Python, R, Java

- **Frameworks:** Pytorch,keras

- **Tools for NLP/ML:** Huggingface-transformers, Sklearn, NLTK, Pandas, Numpy

- **Others and Soft Skills:** LaTeX, Markdown, Linux, Shell, Anaconda