

# Xinyu Ma

---

## CONTACT INFORMATION

Tel: 13611248748  
Email: xinyuma2016@gmail.com  
Home: <https://albert-ma.github.io>

## BIOGRAPHY

Xinyu Ma received the B.E. degree from Shandong University in 2016. He is currently pursuing the Ph.D. degree with the CAS Key Lab of Web Data Science and Technology, Institute of Computing Technology, Chinese Academy of Sciences under supervision of Prof. Jiafeng Guo.

His research focuses on the core problem in information retrieval (IR), i.e., Ranking. Specifically, his current research interests are in leveraging pre-trained language models into IR, including but not limited to understanding the relevance in IR, pre-training for IR and dense retrieval.

He has published several papers on the top-tier conference and journals in IR, such as SIGIR, WWW, WSDM and F<sub>n</sub>TIR. He proposed **the world's earliest** pre-trained model tailored for re-ranking, i.e., PROP, which reached the top of the **world's largest** deep learning ranking task, i.e., MS MARCO Document Ranking Leaderboard.

## EDUCATION

**Chinese Academy of Sciences**, Beijing 2017.9 - 2023.1

Ph.D. Student, Institute of Computing Technology  
Advisor: Jiafeng Guo

**Shandong University**, Jinan, Shandong 2013.9 - 2017.7

B.S., Software Engineering  
GPA Top 2%

## INTERNSHIP

### Sina Weibo, China

Intern in Machine Learning Group, AutoML Project 2018.03 - 2018.09

- Network Structure Generation. A sequence-to-sequence model is used to generate the CNN architecture strings.
- Reinforcement Learning. The generation model is optimized via policy gradient learning according to its performance on CIFAR-10.

### Ant Group, China

(Ongoing) Intern in Search Group, Retrieval at E-commerce Project 2022.8 - Present

- Dense Retrieval Model on Structure Data

## COMPETITION

**1st Place on MS MARCO Document Ranking Task** 2021.1, 2021.4

- We outperform THU-MSR, CMU, Stanford and other teams from different companies and universities, and reached the 1st place twice using PROP
- We are the first team to edge past eval MRR@100 of 0.4.

## HONORS AND AWARDS

ICT President Award (Special Prize) (Highest award in ICT, CAS), 2021  
National Scholarship (Ph.D), 2021  
SIGIR 2021 and 2022 Student Travel Grant  
Outstanding Undergraduate Award from Shandong Province, 2017  
National Scholarship (B.E.), 2016

National Encouragement Scholarship, 2015

## PROJECTS

**CoreNLU**, ICT, CAS, 2018.7-2019.1

Participant, *Natural Language Understanding Engine Project*

- Develop sequence tagging models and System backend
- Contribute 1w+ lines of code

**MatchZoo**, Open Source Project on Github

Participant, *Natural Language Understanding Engine Project*

- Develop text matching models and collect paper resources
- Receive a total of 4K+ stars

## PUBLICATIONS

- [1] PROP: Pre-training with Representative Words Prediction for Ad-hoc Retrieval  
**Xinyu Ma**, Jiafeng Guo, Ruqing Zhang, Yixing Fan, Xiang Ji, Xueqi Cheng.  
The 14th ACM International Conference on Web Search and Data Mining (**WSDM**), 2021, **CCF-B**, Oral
- [2] B-PROP: Bootstrapped Pre-training with Representative Words Prediction for Ad-hoc Retrieval  
**Xinyu Ma**, Jiafeng Guo, Ruqing Zhang, Yixing Fan, Yingyan Li, Xueqi Cheng.  
The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (**SIGIR**), 2021, **CCF-A**
- [3] Pre-train a Discriminative Text Encoder for Dense Retrieval via Contrastive Span Prediction  
**Xinyu Ma**, Jiafeng Guo, Ruqing Zhang, Yixing Fan, Xueqi Cheng.  
The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (**SIGIR**), 2022, **CCF-A**
- [4] A Linguistic Study on Relevance Modeling in Information Retrieval  
Yixin Fan, Jiafeng Guo, **Xinyu Ma**, Ruqing Zhang, Yanyan Lan, Xueqi Cheng.  
The Web Conference 2021 (**WWW**), 2021, **CCF-A**
- [5] Scattered or Connected? An Optimized Parameter-efficient Tuning Approach for Information Retrieval  
**Xinyu Ma**, Jiafeng Guo, Ruqing Zhang, Yixing Fan, Xueqi Cheng.  
The 31st ACM International Conference on Information and Knowledge Management (**CIKM**), 2022, **CCF-B**
- [6] A Contrastive Pre-training Approach to Discriminative Autoencoder for Dense Retrieval  
**Xinyu Ma**, Ruqing Zhang, Jiafeng Guo, Yixing Fan, Xueqi Cheng.  
The 31st ACM International Conference on Information and Knowledge Management (**CIKM**), 2022, **CCF-B**
- [7] An Empirical Investigation of Generalization and Transfer in Short Text Matching  
**Xinyu Ma**, Yixing Fan, Jiafeng Guo, Ruqing Zhang, Lixin Su, Xueqi Cheng.  
Journal of Computer Research and Development, 2021, **CCF-A**, Chinese Journal
- [8] Pre-training Methods in Information Retrieval  
Yixing Fan\*, Xiaohui Xie\*, Yinqiong Cai, Jia Chen, **Xinyu Ma**, Xiangsheng Li, Ruqing Zhang, Jiafeng Guo.  
Foundations and Trends in Information Retrieval (**FnTIR**), IF 8