

Yuze Gao

✉ yuze.gao@outlook.com | ☎ +65 90553268 | 🌐 github.com/CooDL
📍 1 Fusionopolis Way, Singapore

[Education]

Northeastern University

Shenyang, China

B.ENG IN COMPUTER SCIENCE AND TECHNOLOGY

Sep. 2010 – Jun. 2014

Thesis: The Design and Implementation of a Simple Teaching Compiler for C

[Research Experience]

Aural & Language Intelligence (ALI) unit, The Institute for Infocomm Research (I2R), Agency for Science, Technology and Research (A*STAR)

1 Fusionopolis Way, Singapore

RESEARCH ENGINEER

Mar. 2019 – Present

Focus on: Deep Learning for Natural Language Processing

Natural Language Processing Lab, Singapore University of Technology and Design (SUTD)

Changi, Singapore

RESEARCH ASSISTANT

Sep. 2016 – Aug. 2018

Focus on: Deep Learning for Natural Language Processing

Natural Language Processing Lab, Northeastern University

Shenyang, China

RESEARCH ASSISTANT

Jul. 2014 – Apr. 2016

Focus on: (Hierarchical) Phrase-based Statistical Machine Translation

Yatrans Network Technology Co., Ltd.

Shenyang, China

RESEARCH ENGINEER

Sept. 2015 – Mar. 2016

Focus on: Frame-based Statistical Machine Translation

[Publications]

2019:

Feature-Less End-to-End Nested Term Extraction

Yuze Gao, Yu Yuan

In Proceedings of the Natural Language Processing and Chinese Computing (NLPCC), Explainable Artificial Intelligence (XAI), 2019

2018:

Learning How to Self-Learn: Enhancing Self-Training Using Neural Reinforcement Learning

Chenhua Chen, Yue Zhang and Yuze Gao

In Proceedings of the International Conference on Asian Language Processing (IALP), 2018

Cross-lingual Terminology Extraction for Translation Quality Estimation

Yu Yuan, Yuze Gao, Yue Zhang and Serge Sharoff

In Proceedings of 11th edition of the Language Resources and Evaluation Conference (LREC), 2018

2017:

Implicit Syntactic Features for Targeted Sentiment Analysis

Yuze Gao, Yue Zhang and Tong Xiao

In Proceedings of the 8th International Joint Conference on Natural Language Processing (IJCNLP), 2017

2015:

A Comparison of Pruning Methods for CYK-based Decoding in Machine Translation

Yuze Gao and Tong Xiao

In Proceedings of the 11th China Workshop on Machine Translation, 2015

[Patents]

2016:

Syntactic Skeleton based Statistical Machine Translation System

CN201610053560.2, Yatrans Network Technology Co., Ltd

Co-inventor, with Tong Xiao, Jingbo Zhu and Chunliang Zhang

[Academic Service]

2018:

Second Reviewer of ICPCSEE 2018, **Second Reviewer** of NAACL-HLT 2018, **Second Reviewer** of EMNLP 2018

2017:

Program Committee of The 9th SIGHAN Workshop on Chinese Language Processing

2015:

Publishing Editor of the Proceedings of the 11th China Workshop on Machine Translation (CWMT 2015)

[Research Projects]

1. CHINA CAPITAL MARKET AND UNSTRUCTURED DATA (CCMUD)

Singapore University of Technology and Design *May 2018 - Present*

CCMUD is a stock related program for market predication. The project consists of a series of NLP tasks, such as segmentation, POS-tagging, named entity recognition, event detection, and sentiment analysis. Most of the tasks are implemented using deep learning frameworks. Specifically, I have been working on segmentation, named entity recognition and sentiment analysis.

2. CROSS-FUNCTIONAL INFORMATION SYSTEMS FOR DECISION MAKING (CISDEM)

Singapore University of Technology and Design *Oct. 2016 - Jun. 2018*

CISDeM is a cyber-information research program for decision making against network security threats. This project adopts an inter-disciplinary research approach – spanning security, natural language processing, machine learning, signal processing, and networking – to protect mission-critical network infrastructures. I have been contributing to the natural language processing part. In particular, I have been working on the detection and clustering of events from the streaming media using deep learning methods. These events are then further analyzed to see whether they would pose a potential threat to the network infrastructure so that early countermeasures can be taken.

3. IMPROVING (HIERARCHICAL) PHRASE- AND FRAME- BASED STATISTICAL MACHINE TRANSLATION

Northeastern University *Aug. 2014 - Mar. 2016*

This project aims to speed up and enhance the decoding process in the (hierarchical) phrase- based and frame-based statistical machine translation system while maintaining its BLUE, which is an index of the translation accuracy. One solution is to apply pruning in the decoding process. Since natural language translation systems share some principles with compilers (rule-based translation system), thus we can use some similar hard or soft constraints, such as derivation rules or grammatical structures, to direct its decoding progress. In particular, I successfully applied syntactic parse trees and boundary information into the optimization of pruning in the decoding process of natural languages. This gained significant decoding speed-up with acceptable loss of translation accuracy.

4. THE DESIGN AND IMPLEMENTATION OF A SIMPLE TEACHING COMPILER FOR C

(Thesis Project)

Northeastern University *Mar. 2014 - Jun. 2014*

This project aims to develop a teaching compiler for the course “Compiler Theory.” The envisaged tool aims to demonstrate the compilation process clearly by giving the detailed intermediate result of each stage. Current implementation supports simple C constructs (like while, for, and if) and data types (like int, float, and array). The supported intermediate results include lexical analysis results (token sequence, lexical analysis errors and identifier list), syntax analysis (first set, follow set, select set, terminal symbol, non-terminal symbol, analysis steps and forecasting analysis table), and translation results (intermediate language like quart expression, symbol table and the target assembly language).

[Internships]

INSTITUTE FOR INFORMATION MANAGEMENT AND INTEGRATION

Northeastern University

Shenyang, China

Dec. 2013 - Feb. 2014

Project: A Programming Experience Sharing Forum

Responsibility: The goal of this project is to develop a web-based programming experience sharing platform using the SSH (Struts, Spring, Hibernate) framework for Java web applications. As leader of a four-person student team, apart from adjusting the progress of the project, I was also responsible for the design and implementation of the database layer of the forum.

INSTITUTE FOR COMPUTER TECHNOLOGY APPLICATION

Northeastern University

Shenyang, China

Apr. 2013 - May 2013

Project: A Web-based Management System for Motorcycle Shops

Responsibility: The project aims to build a web-based management system for a motorcycle company with many nation-wide branches. In particular, I was responsible for the frontend UI design.

[Technical Skills]

NATURAL LANGUAGE PROCESSING:

- Proficient with Statistical and Syntax-based Machine Translation, Parsing, Sentiment Analysis and Deep Learning for NLP
- Interested in **Opinion and Sentiment Mining, Machine Translation and Language Parsing**

DEEP LEARNING FRAMEWORK:

- Proficient with **Tensorflow, PyTorch**.

COMPILERS:

- Familiar with compiler concepts and programming language processing.
- Have some experience in applying compiler techniques to Statistical Machine Translation Systems

PROGRAMMING LANGUAGES:

- **Major:** C++, Python
- **Minor:** Perl, Assembly Language

PROGRAMMING TOOLKITS:

- Proficient with Linux Programming Environment.
- **Tools:** GCC, GDB, GNU Make, Eclipse, Pycharm