YUAN XIN

Address: 9#320, Female Dormitory, University of Science and Technology of China, Hefei, Anhui
Email: lwjxy@mail.ustc.edu.cn Tel: (+86) 188-5608-3742

EDUCATION

UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA (USTC) Hefei, China
Master in Computer Science ; Major score: 3.07/4.3 09/2018-06/2021

UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA (USTC) Hefei, China
Bachelor in Computer Science ; Major GPA: 3.18/4.3 09/2014-06/2018

MONASH UNIVERSITY (***exchange student***) Melbourne,Australia

Graduate project: Satellite Image Classification Based on LSTM 06/2018-03/2018

INTERNSHIP

BAIDU ***NLP Engineer Intern***, ***Baidu Talent Intelligence Center*** Beijing, China 06/2020–09/2020

Project1: Extract hierarchical relations between skills in the JD dataset.

- Preprocess and clean text data, transform the data into the proper training set and test set. To ensure the proportion of training and test set, use sampling to control the amount of test set.
- Use three encoding methods(general encoding, pattern encoding, and position encoding) to represent the sentence
- Feed the representation of sentence into CNN/BiLSTM
- Make binary classification for samples and achieved a 94.6% F1-score.
- Use an iteration method to improve classification accuracy thus to extract more relation pairs

Project 2: Skills representation

- Preprocess data (a sentence contains several skills and we need to learn representations for all skills in the sentence)
- View a skill as a sentence, tokenize it and use CNN/BiLSTM to encode skills, then send the encoded skills into word2vec (cbow+negative sampling) to learn its representations
- View a skill as a sentence and use sentence-bert to learn representations. Specifically, use negative sampling to sample the negatives, then use the triple format objective to learn its representation.

ALIBABA DAMO ***NLP research Intern*** Hangzhou, China, 09/2021 – now

Doing research on machine translation and self-traning

PROJECTS

GCN for text classification. 10/2019-12/2019

1) Label-Incorporated GCN

- Build a heterogeneous graph which contains document node, word node, label node
- Use a two-layer GCN to learn node representations
- Transform document classification problem into node classification problem and performs better than Text-GCN(AAAI'19) in three datasets (WebKB, Ohsumed, R8) for about one percent.

2) Heterogeneous GCN

- Divide the heterogeneous graph into different sub-graphs, each subgraph represents a kind of relationship
- Use different GCNs to deal with the sub-graphs and this achieves competitive results against the first one
- This method is inductive and has low computational complexity

PUBLICATION

Label Incorporated Graph Neural Networks for Text Classification ***ICPR'2020***

Yuan Xin, Linli Xu, Junliang Guo, Jiquan Li, Xin Sheng

SKILLS

- C,Python, Pytorch,Keras

HONORS & AWARDS

Outstanding Student Scholarship, USTC 2016

SCORES OF MAJOR COURSES

Algorithm Desigh and Analysis:87.6
Advanced Artificial Intelligence:87
Function of Complex Variable:86
Operating System:89
Computer Organization:87
Combinatorial Mathematics:93
Graph Theory:85
Calculus For Functions of a Single Variable:85
Mathematical Logic:87
Software Engineering:86