Haoxuan (Horace) Wang

+44 (0)7774857427

billweasley20092@gmail.com

Github: https://github.com/billweasley

Linkedin: https://www.linkedin.com/in/horace-haoxuan-wang

Education

- 2018 - 2019 (Expected) **University College London**, Msc Web Science and Big Data Analytics

  Core subjects: Probability Graphical Models; Introduction to Deep Learning; Complex Network; Affective Computing; Statistical NLP; Information Retrieval; Multi-agent AI, Applied Machine Learning

- 2016 - 2018 **University of Liverpool**, Bsc Internet Computing, 1st Class with an average of 80.75 %

  Core subjects: Software Engineering; Database Concepts; Internet Principles (Introduce to OSI layers); Object-Oriented Programming; Distributed Systems Concepts; Software Development Tools (Mainly about testing); Principles of C && Memory Management; iOS Programming (Swift); Knowledge Representation & Reasoning; E-commerce (Auction and Security [RSA, Diffie–Hellman key exchange, Elliptic Curve Encryption]);

- 2014 - 2016 **Xi'an Jiaotong-Liverpool University**, Bsc Information and Computing Science

  2+2 pathway routine (first 2 years in Suzhou, China and final 2 years in Liverpool, UK), dual degree.

  Core subjects: Computer Systems; Introduction to Databases; Introduction to Programming in Java; Algorithmic Foundations and Problem Solving; Data Structures; Operating Systems Concepts; Human-Centric Computing; Calculus; Introduce to Discrete Mathematics

Work Experiences

- 2019.08 - Now **Barclays PLC UK** Technology Graduate

  1st Rotation: Barclaycard Backend Shared Service Web API Development

Projects

- 2019.06 - 2019.09 **Project Internship (Master Degree Thesis) @ Astroscreen**

  Social media posting language source identification (tweets and gabs) project. Finished a crawler for collecting language (posts) data from Gab.com, pre-processed data using Regular Expression, built models for classifying the source of these data by fine-turning BERT and XLNet, visualised results using t-SNE, did "leave-one-hashtag-out" cross-validation and evaluated models using some common matrics (Accuracy, F1 score, Confusion Matrix, Matthews Correlation Coefficient).

  - Python, Karas, Tensorflow, MulticoreTSNE, Matlabplot

- 2019.03 - 2019.04 **Information Retrieval Course Project**

  Multiple practices using Fact Extraction and Verification (FEVER) dataset。
  Including word counting and verfication of zip's law; implementation of vector space information retreive (TF-IDF); inplementation of query likelihood document retreive (applying Laplace Smoothing , Jelinek-Mercer Smoothing and Dirichlet Smoothing, respectively); implementation logistic regression to predict senatance relevance; implementation of Precious, Recall and F score function; using nenral network to predict document truthness.

- 2019.02 - 2019.03 **Integrating BERT and Embeddings into CommonsenseQA Chanllenge**

  We fine-turned Google BERT to CommonsenseQA chanllenge 1.0 (with 3 options of each question) and then integrated Conceptnet Numberbatch and ELMo embeddings attempting to improve the model proformance. The chanllenge involves a set of MCQ questions that requiring human commonsense knowledge. We achieved 68.79% of accuracy on validation set using BERT + ELMo (soly BERT : 67.47%; BERT + Numberbatch: 67.68%).

  - Python, Keras

- 2019.02 - 2019.02 **"Recogising food-stimalated emotions" experimental labelling platform**

  A simple web app that we used to collect experimental data (food taste affection). An online demo could be visited at https://affective-computing-data-collection-dist.cfapps.io/sessions （Offine now, because out of charge）. I finished this on own on a weekend. This is a draft version of it (we then removed the personal info and produced a local version of it).

  - Bootstrap, JQuery, Recordrtc.js, Java, Play! framework 2, PostgreSQL, Cloud Foundry

- 2019.02 - 2019.03 **Maximise number of clicks through AD CTR prediction and bidding functions selection**

  Pedicting whether user would click the online AD (advertisement) on a AD real-time DSP bidding history dataset. The prediction results then were inputted to a bidding strategy function to predict a bid price. The total pay price is bounded by a constant total number. The dataset is unbalanced with only about 3000 positive samples (clicks) among more than 300000 bidding records. We tried many different different models (XGBoosting, Shallow NN, Logistic Regression) and some bidding stratgies. We also applied downsampling and re-calibration techniques in the project. We did a competition in a leaderboard with other students (30 groups) and ranked a 3rd place (with 185 clicks and the first 2 are with 186 clicks).

  - Python, Karas, XGBoost, Numpy, Pandas, Matlabplot

- 2018.12 - 2019.01 **Is a uploader with more uploaded videos also more popular? A network based analysis on bilibili**

  Bilibili is one of largest Chinese video sharing website. The project aims to examine some properties (degree distribution and assortative coefficient) of the user (uploader) following relation network and then attempts to check if they are related between the number of archives (that reflecting how the user active is) and the in-coming degree of the nodes (that reflecting how popular the user is) through visualisation of the network. A crawler was written in the project to accuire data from Bilibili's RESTful API.

  - Python, networkx, graph-tool, MySQL

- 2017.09 - 2018.05 **Simulation, Visualization and Experimental Analysis for Population Protocols and Network Constructor in 2-Dimension** *(https://github.com/billweasley/Bsc-dissertation)*

  Population protocol is a theoretical model for distributed computation. The model contains a collection of indistinguishable agents. The network constructor and the terminating grid network constructor are some models extending population protocol but with a different aim to construct network in different topologies.

  - Kotlin, GraphStream