

# Yuhan Wei

Tel: +852 65001654 Email: yuhan\_wei@whu.edu.cn

## EDUCATION

**Wuhan University**, Academic Master's Degree in Data Science

Wuhan, China

- Accumulative GPA: 3.53/4.00 Sep 2019-Jun 2022
- Member of Information Retrieval and Mining Institute of Wuhan University
- Core Courses: Information System Engineering, Natural Language Processing, Principles of Machine Learning, Data Visualization, Mathematical Modeling
- Honors: Second-class Scholarship for Academic Excellence (TOP 2/8, 2020); Excellent Student Leader (2020)

**Nanchang University**, B.S. in Information Management and Systems

Nanchang, China

- Accumulative GPA: 3.61/4.00; Class Ranking: **1/39** Sep 2015-Jun 2019
- Outstanding Student Scholarship (TOP 2/40, 2016/2017/2018); Merit Student & Excellent Student Leader (2016 & 2019)

## PAPER & PATENT

- **Yuhan Wei**, Xiaohui Vincent Wang. (2023). "The Role of Online News Media in the Global Media Ecosystem: Analyzing the Multimodal Network of International News Flow", *73<sup>rd</sup> Annual ICA Conference*, Accepted
- **Yuhan Wei**, Wei Lu, Qikai Cheng, Tingting Jiang, Shewei Liu. (2022). "How Humans Obtain Information from AI: Categorizing User Messages in Human-AI Collaborative Conversations", *Information Processing and Management*, 59(2), 102838. **JCR Q1**
- Tingting Jiang, Qian Guo, **Yuhan Wei**, Wei Lu, Qikai Cheng. (2022). "Investigating the Relationships Between Dialog Patterns and User Satisfaction in Customer Service Chat Systems Based on Chat Log Analysis", *Journal of Information Science*, 01655515221124066. **JCR Q1**
- Jinjing Yang, Shengzhi Huang, **Yuhan Wei**, Zhifeng Liu, Wei Lu. (2020) "An Improved Framework for Identifying Emerging Author Keywords", *Proceedings of the 10th Global TechMining Conference*, November 11, 2020 Virtual Event (Power Talk)
- Jianqing Yang, **Yuhan Wei**, Shengzhi Huang, Wei Luo, Wei Lu. (2020). "Review on Emerging Topic Recognition Based on Scientific and Technological Literature", *Information Science*, (08), 159-163. CSSCI
- Wei Lu, Jinjing Yang, **Yuhan Wei**, Leyan Wu. "Science and Technology Information Analysis Combined with Time Lag Calculation in the Subject Evolutionary Reasoning Method", *National Invention Patent*, 201911081064.8
- Huaping Gong, **Yuhan Wei**, Jialiang Liu. (2018). "The Role of Competitive Intelligence in the Process of Scientific and Technological Innovation—A Case Study of SMEs", *Modern Information*, 39(01), 118-125. CSSCI

## RESEARCH

**Detecting and Predicting International Financial Relations**

Jul 2020-Present

*Laboratory for AI-Powered Financial Technologies Limited*

*Project Leader: Dr. Xiaohui Vincent Wang*

- Established a database of 4235037 pieces of conversation data about insurance consulting, applied MySQL to process and clean the data, and used natural language processing to normalize conversations with Python, including word segmentation, part-of-speech tagging and key-phrase extraction.

**How Humans Obtain Information from AI: Categorizing User Messages in Human-AI Collaborative Conversations**

Jun 2020-Mar 2021

*Information Retrieval and Mining Institute of Wuhan University*

*Collaborated by Intelligent Insurance Lab of Tk.cn*

*Project Leader, Supervisor: Prof. Wei Lu*

- Established a database of 4235037 pieces of conversation data about insurance consulting, applied MySQL to process and clean the data, and used natural language processing to normalize conversations with Python, including word segmentation, part-of-speech tagging and key-phrase extraction.
- Divided conversation into three types of user-AI, user-human, and mixed representatives, and concluded a categorization system for user messages based on the message intention; conducted data annotation experiment; summarized and reported the characteristics of different message types and compared their usage in sessions; visualized the experiment results by using Echarts.
- Plan to study the structure of user intent in conversation and achieve the automatic prediction of user intent structure by deep-learning models, Logistic Regression and SVM have been applied already.

**User Information Behavior in Dialogues between Users and Intelligent Customer Service**

Apr 2020-Dec 2020

*Center for Studies of Information Resources of Wuhan University*

*Supervisor: Prof. Wei Lu/Prof. Tingting Jiang*

- This study represents an obtrusive chat log analysis that followed the established approaches of search log analysis and examined the relationships between dialog patterns and user satisfaction.
- Used Python to process and classify 251556 pieces of user messages from WeSure, an insurance agency platform and

# Yuhan Wei

Tel: +852 65001654 Email: yuhan\_wei@whu.edu.cn

selected out 75918 rounds of question dialog.

- Processed data cleaning, merging and coding, experimented with the above dialog act categorization and matching rules on a random sample of 5,000 user messages extracted from the log.
- Extracted 147,972 sub-sessions and analyzed in terms of topic, length, and path; found that the chatbot may be responsible for damaging user satisfaction in some cases.

## Evolution of Research Topics and Identification of Emerging Keywords

Sep 2019-Dec 2020

Information Retrieval and Mining Institute of Wuhan University

Affiliated with "Literature Citation Recommendation Based on Multi-semantic Information Fusion" Project

Sponsored by National Natural Science Foundation of China, Supervisor: Prof. Wei Lu

- Proposed a dynamic novelty indicator method (DNIM) with a nested attenuation factor to measure the dynamic novelty values of different topics based on literature review.
- Programed a web crawler with Python, obtained metadata of 157763 papers from ACM database from 1951 to 2018 which contained 219903 keywords.
- Used Natural Language Toolkit to normalize the keywords, fitted the Top 500 keywords regarding use frequency by adopting logistic regression model, and validated the method by comparing the computation result of top 0.05%, 0.1%, 0.5%, 1%, and 1.5% keywords with benchmark model and DNIM.

## COURSE PROJECTS

### Analysis and Differtermination Model Based on Machine Learning

Jan 2021-Sep 2020

- Proposed a P300 event identification model based on Linear Discriminant Analysis that achieved an accuracy of 72.8% on the test data of five subjects, took a series of brain telecom signal preprocessing of filtering, reduced sampling and normalization.
- Proposed a sampling channel scoring model based on Independent Component Analysis, which scores various channels by demixing matrix between sampling channels and independent components, the method achieved an average accuracy of 76%.
- Used a graph construction algorithm based on data self-representation and Laplace smoothing constraints, and a SRLS graph-based label transfer algorithm for semi-supervised learning of P300 signals.
- Applied decision tree algorithm and multi-layer percepmachine model to learn sleep data, with 70.4% and 74.8% accuracy on the test sets respectively.

### Model-based Global Language Trends Analysis

Mar 2020-June 2020

- Obtained and computed the number of native speakers from 2014 to 2016 and data of immigration rate, population growth rate, GDP, and cultural power for the thirteen language-speaking countries.
- Built population growth and improved-Markov models to predict population change and geographical distribution of these languages in the next five years.

### Forecasting of Agricultural Hot Topics Based on Multi-source Data Analytics

Sep 2019-Jan 2020

- Extracted 22657 pieces of fund project texts, 163778 patent documents, 135790 journal papers, and 15565 conference papers in agricultural area from 2008 to 2014.
- Divided the abstract texts into 32 datasets, processed word segmentation, word stemming, transformed the data into sparse vector, extracted topics through Latent Dirichlet Allocation, and visualized the relations among hot topics with Matplotlib.

## DIVERSITY & LEADERSHIP

Teaching Assistant for Undergraduate Class "Classic Guide to Humanities and Social Science	Sep 2020-June 2021
Leader of Class Arts Community	Sep 2019-June 2021
Volunteer at ACM/ IEEE Joint Conference on Digital Libraries 2020	Aug 2020
Assistant to Student Advisor of School of Management	Sep 2018-June 2019
Participant in the 2018 Global Management Challenge (First Prize)	Jan 2019
The 17 <sup>th</sup> National Enterprise Competition For MBA Training Colleges (Second Prize)	Apr 2018
Honorable Mention for MCM: The Mathematical Contest in Modeling	Feb 2018
Participant in the 2017 Global Management Challenge (Second Prize)	Jan 2018
Leader in the 10 <sup>th</sup> National College Students Network Business Innovation and Application Competition (Outstanding Winner)	Dec 2017
Leader of Student Learning Service Center	Sep 2016-June 2017

## SKILLS

- Computer Skills: Python, Java, SQL, HTML, SPSS, C++
- Language Proficiency: Native in Chinese, Fluent in English ( IELTS 7.5)