

# Cloud Computing and Big Data

Dr. Na Helian ([n.helian@herts.ac.uk](mailto:n.helian@herts.ac.uk))

I have been supervising Ph.D projects in the following topics.

1. Customer churn prediction based on data mining/machine learning techniques
2. Accurate and efficient bacterial colony counting by using deep learning
3. Feature weighting and instance weighting unsupervised learning
4. Assessing variability of EEG and ECG signals using a variety of non-linear measures
5. Optimising hierarchical load balancing for cloud computing

We are looking for a PhD student who would be interested in the areas of Cloud Computing and/or Big Data analysis.

The prospective candidates should have a strong background in Computer Science. In particular, they should demonstrate strong programming skills in one or more major computer languages.

Note: The topic “Cloud Computing and Big Data” has to be given on the application form.

Please contact Dr. Na Helian ([n.helian@herts.ac.uk](mailto:n.helian@herts.ac.uk)) directly if you have any specific questions on those research areas.

## Ph.D. Project Ideas

### 1. Twitter Tweets sentiment Analysis using deep learning

Key words: Deep Learning, text mining, Natural Language Processing

People on Twitter post messages in real time about their opinions on a range of topics. Therefore, Twitter is a gold mine of data for Sentiment Analysis. Substantial research has been conducted on Twitter Tweets sentiment analysis using various types of Machine Learning Algorithms, including supervised, unsupervised. But there is not much work done by using deep learning although deep learning has been extensively applied in a lot of other applications. The aim of this project is to use deep learning in sentiment analysis on Twitter tweets.

### 2. Using deep learning for user authentication by keystroke dynamics

Key words: Deep Learning, User Authentication, Keystroke Dynamics

User identification is the process of deciding whether someone is in fact whom he/she claims to be. The wide use of online services requires reliable, widely available and cheap person identification techniques. The science and technology of identification can use individual’s physiological or behavioral biometrics to identify them. Keystroke dynamics is a behavioral biometrics and it uses the rhythm and pattern which each individual types. The typing manner is characteristic to individuals, and a user is hardly able to copy the typing dynamics of other users. In the previous works and researches done in this area, user typing patterns or timing vectors of their typing are measured and used in various machine learning algorithms for user authentication. But there is not much research conducted in the use of deep learning and neural network

based approach. The aim of this project is to use deep learning and deep neural network in keystroke dynamics for user authentication.

### **3. Optimising hierarchical load balancing for cloud computing**

Key words: Clouding Computing, Scheduling, Hierarchical Load Balancing

Cloud Computing plays a major role in distributed systems. Cloud provides services which benefit us in many ways, such as reliability, robustness, quality of service. Job scheduling is a method to distribute workload across multiple resources (e.g. computers) over cloud to achieve optimal resource utilization. There are many ingenious load balancing algorithms. But there is not much research conducted in scheduling structure design. Hierarchical load balancing uses multiple layers of load balancing algorithms in a tree structure. The proposed research is to enhance Data Centre efficiency by studying the features, benefits and by the plan of hierarchical load balancing topologies. The aim of this project is to optimize the designed structure using artificial intelligence technologies.

The following projects are part of a larger project for a local company. The overarching project aim is to create an Adaptive Virtual Agent which is capable of performing debt recovery.

### **4. Feature extraction from speech to text for big data**

Key words: Big Data, Natural Language Processing

The ever decreasing cost of digital storage has enabled the storage of vast amounts of audio recordings of telephone calls. Combining this with a drive for automation in industry, the importance of processing audio data into more usable text, and ultimately extracting features from the data is clear. The aim of this project is to evaluate/design state of the art technologies in Speech to Text for Big Data and feature extraction. This project is straight from industry and presents a cutting edge real world problem. The research outcomes will be deployed in industry, which should provide a wealth of feedback.

### **5. Automatically generating conversational trees**

Key words: Data Mining, Natural Language Processing

The use of chatbots is on the rise. However implementing conversational agents such as chatbots can be labour intensive and less comprehensive. The aim of this project is to establish novel solutions for building a tree/graph from millions of conversations and other data. The outcomes of project will guide the creation of chatbots with a much higher capability/intelligence level than those currently found in consumer devices. This project is straight from industry and presents a cutting edge real world problem. The research outcomes will be deployed in industry and this will provide a wealth of feedback.