# DONGKAI CHEN

**Email:** Dongkai.Chen.GR@dartmouth.edu **Github:** https://github.com/HOLMEScdk
**Add.:** 8 Sachem Cir, Apt 6, West Lebanon, NH, 03784 **Tel:** +1 6175952078

## EDUCATION

**Dartmouth College, USA** *Sep. 2019 - Jun. 2021(Expected)*
Master of Science, Department of Computer Science Advisor: Prof. V.S. Subrahmania

**Zhejiang University City College, China** *Sep. 2015 - Jun. 2019*
Bachelor of Engineering, Department of Computer Science GPA: 3.95/4.0, Rank: 1/67

## PUBLICATIONS

- Canghong Jin, Dongkai Chen, Fanwei Zhu, and Minghui Wu. Detecting suspects by large-scale trajectory patterns in the city. *Mobile Information Systems*, 2019:1837594:1–1837594:11, 2019
- Canghong Jin, Haoqiang Liang, Dongkai Chen, Zhiwei Lin, and Minghui Wu. Identifying mobility of drug addicts with multilevel spatial-temporal convolutional neural network. In *Advances in Knowledge Discovery and Data Mining - 23rd Pacific-Asia Conference, PAKDD 2019, Macau, China, April 14-17, 2019, Proceedings, Part I*, pages 477–488, 2019

## PROJECTS

**Preserving Styles in Text Generation** (Ongoing)

- Used the estimated stylistic property distributions as the prior of the structured latent code and introduced Transformers as Encoder/Decoder structure.
- Developed a separate fine-tuning phase for the decoder that incorporates two extra regularization terms in the decoder loss function to penalize the differences in styles of the genuine and synthetic samples.

**Understand Mathematical Questions with Hybrid Knowledge and Pretrained Language Model**

- Designed a framework which combines content tokens and mathematical knowledge concepts in whole procedures. 1)Embedded entities from mathematics knowledge graphs, 2)integrated entities into tokens in masked language model, 3)introduced semantic similarity-based task as next sentence prediction, and 4)fused knowledge vector and token vector during the fine-tuning procedure.
- The performance of KG-MTP model outperforms other state-of-the-art model, like BERT and ERINE.

**Mining spatio-temporal data to classify pedestrian trajectories**

- Conducted data cleaning and integration of the spatio-temporal data of existing big data set.
- Established a trajectory classification model combined with features extracted from a graph model(inspired by GCN) and CNNs to identify trajectories belong to citizens or criminal suspects.

**Search engine based on Chinese Quroa users (million level)**

- This project includes designing crawler rules(*a simple distributed architecture*), maintaining databases(*Redis MongoDB*), indexing(*ElasticSearch*), segmenting text(*Jieba*), and responding to queries.
- With simple multi-processes and multi-thread processing, the amount of regular inbound data could reach 20W per day, and the query speed of ElasticSearch is about 0.1s.

## INTERNSHIP

Algorithm Engineer at EvaVisdom (11.2018-03.2019)

- Worked on deep learning in computer vision and post-processing after model prediction to maintain a robust model in real scenario.
- Under the supervision of experts in this field from NLPR, and did the research topics on data mining and time series for dirty data and anomaly detection.

## AWARDS AND HONORS

- **2017-2018** *42th ACM International Collegiate Programming Contest Asia Regional* ***Silver***
- **2017-2018** *42th ACM International Collegiate Programming Contest Asia Regional* ***Bronze***
- **2016-2017** *China Undergraduate Mathematical Contest in Modelling* ***National Second Prize***
- **2016-2017** *Mathematical Contest in Modelling (MCM/ICM)* ***Honorable Mention***
- **2016-2017** ***China National Scholarship***

## SKILLS

- **Programming Language:**
  Python(Proficient), C/C++(Competent), Java(Competent)
- **Domain Knowledge:**
  Data Mining, Deep Learning, Machine Learning