

Education

2014.09-now	Dalian University of Technology Top 1 of Recommendation	Data Mining	Master
2010.09-2014.07	Dalian University of Technology Top 3.3%	Software Engineering	Bachelor

Skills

-
- ♦ Language: C++, Python
 - ♦ Algorithm: Frequent-used fundamental algorithms
 - ♦ Machine learning: Popular algorithms of classifier, clustering, association rules, recommendation, topic model
-

Research and Project Experience**2014.11-2015.07 Conditional Discriminative Pattern Mining: Concepts and Algorithms**

- ♦ Most existing approaches to discovery discriminative patterns usually generate many redundant patterns. To remove the redundancy, we propose an effective algorithm called CDPM to generate a set of non-redundant discriminative patterns. We have submitted our paper to the journal of **Information Science (CCF B, SCI IF = 3.364)**. The paper has been revised twice.
- ♦ In this project, I designed the algorithm, did the experiments and wrote the paper.

2014.03-2012.07 Statistical evaluation of the clustering results based on permutation test

- ♦ For the result of clustering, we propose a new algorithm called ECP, which uses permutation test to evaluate the result of clustering. This paper has been published by **CAAI Transactions on Intelligent Systems**.
- ♦ In this project, I mainly designed the algorithm, did the experiments and wrote the paper.

2012.11-2012.12 Chinese Character Recognition

- ♦ First, color-images are preprocessed by graying and binarization. Second, we extract feature vectors from various directions. And then, we utilize KNN to classify the characters.
- ♦ In this project, my contribution is to design the algorithm and write the code in the part of recognition.

Internship Experience**2016.05-now DiDi Research • 7th Strategy Team**

- ♦ Prediction of the number of orders. We split every city into thousands of hexagonal grids whose diameter is 1km. With the historical data, we build 66 features and utilize Linear Regression to predict the number of orders in 10 minutes and the error rate is 22.7%. In this project, I am responsible for data processing, train and optimize the model. The project has been online.

2016.01-2016.05 Amazon • Capacity Plan

- ♦ Economic Operation Plan. The goal of this project is to make a plan for supply chain of the warehouses in America by considering the cost of shipment and labor. In this project, I am responsible for data processing including filling missing data and smoothing outliers. In addition, I develop a tool to evaluate the plan. The project has been online.

2013.10-2014.02 Baidu • ECOM • Anti

- ♦ Manual Evaluation Platform. I participate in developing this project (including “ PC search”, “PC click”, “Mobile search”, “Mobile click”, “Web sites Baidu Union”). Based on Map Reduce, we sampled the logs in Baidu search engine, calculated the corresponding values to evaluate anti-spam strategies and improve the classification accuracy.

2013.06-2013.09 Microsoft • STC • Bing

- ♦ Grab the data from Tmall. For a shop in Tmall, its corresponding data about behaviours of consumers are grabbed every five seconds and then inserted into a database. After analyzing the data, the result is shown in a website with real-time updates. In addition, the data are also utilized by other teams to perform other experiments.

Blog

<http://www.cnblogs.com/gufeyang/>