ZEXI YU

zexi.yu@usask.ca
+1-(306)3619022
/in/zexi-yu-zackcode
ZackCode

Summary

As a data scientist, I specialize in machine learning algorithms related to medical image processing and biometrics. Besides my expertise, I also have working knowledge of other parts of data science, such as data wrangling, inferential statistics, exploratory data analysis, SQL operations, and Spark operations.

Skills

DATA SCIENCE RELATED

Machine Learning
Pattern Recognition
Data Wrangling
Inferential Statistics
Classification Algorithms
Regression Algorithms
Clustering Algorithms
Deep Neural Networks
SQL
Hypothesis Testing

CODING LANGUAGE AND RELATED PACKAGES

Python
Matlab
Pandas
Scikit-learn
Tensorflow
Spark

OTHERS

Medical Image Processing
Bio-signal Processing

Education

Springboard
Certificate Data Science Career Track 2018

University of Saskatchewan
M.Sc Electrical and Computer Engineering 2017

University of Science and Technology of China
B.Sc Electrical Engineering 2012

Employment

University of Saskatchewan — Saskatoon
Research Assistant — 2012 to 2017

I conduct research objectives related to machine learning and pattern recognition, especially on medical imaging (PET, CT, and MRI) and
biosignals (ECG and SCG). Specifically:

- Using auto-correlation and multi-class regression for human identification using Seismocardiography (SCG)
- Using graph-cuts and SVM for accurate boundary segmentation of lung cancer on PET-CT images
- Adapting learning algorithms to large data scenario by dividing the data into chunks and perform chunk-by-chunk learning

Projects

Mini-Projects — Aug 2017 to Jan 2018

Mini projects are a collection of series of practices demonstrating my other aspects of data science skills, highlighting:

- SQL and JSON practices
- MapReduce with Spark practices

Lung Tumor Detection in CT images — Nov 2017 to Jan 2018

The task of this project is detecting lung tumor in CT images. After the data is cleaned, a Convolutional¤ Neural Network is used in this project. With several limitations, the performance beat the competition benchmark on the Kaggle site. The coding language for this project is Python and the learning algorithm is built using Tensorflow.

Walmart Sales Prediction in Extreme Weather — Aug 2017 to Oct 2017

The task of this project is predicting Walmart sales during extreme weather events. There are three main parts of this project:

- Data Wrangling: The datasets are merged, missing values are addressed, outlier search is conducted, and the linear dependent features are excluded.
- Inferential Statistics: Hypothesis techniques, such as A/B testing, are used to find the relationships between features and prediction target. These relationships are later used to infer the appropriate prediction models used in machine learning section.
- Multiple machine learning algorithms, such as SVM, Decision Trees, Random forest, and Neural Networks are used. The performance overall is comparable to a silver entrance on the Kaggle competition site.

Python is the coding language used in this project. Related packages are Pandas, Scikit-learn and Tensorflow.

Masters Thesis: Tumor Recognition and Segmentation in PET-CT images — Mar 2015 to Jul 2016

The task of my thesis is to provide an automatic and accurate segmentation method for lung tumor regions. Graph-cuts is deployed to merge the information from PET and CT, therefore accurate boundary can be drawn for high-radiation regions in low-resolution PET images. After that, multiple image feature extraction methods are used to extract features from the high-radiation regions, and SVM is deployed to identify tumor from those regions. Overall segmentation performance is 93% dice index rate. Coding language for this project is Matlab.

Biometrics using Seismocardiography (SCG) — Oct 2013 to Oct 2014

The task of this project is to identify people using their SCG signals. The autocorrelation method is used to extract features from SCG signals. Then, multi-class regression method is used for the identification task. The overall performance is 95% identification accuracy on over 300 samples from 25 users. Coding language for this project is Matlab.

Awards

Department of ECE, University of Saskatchewan · Electrical & Computer Engineering Devolved Scholarship — 2014

Royal University Hospital · Royal University Hospital Scholarship — 2013

Ministry of Education of China · National Scholarship — 2010