DATA SCIENTIST

Summary

I am a data engineer/analyst with a passion for statistics/machine learning/deep learning and focused on providing insightful data-driven solutions. I have 5+ years experience in data analytics and data engineering in various fields of manufacturer, finance and infrastructure.My current research and development interests include data engineering, machine learning, deep learning, NLP, and experimentation.

Skills

- Data Visualization: D3.js, Tableau, R visualization packages, Microsoft Excel
- Data Analytics Tools/Programming:Python (scikit-learn, numpy, scipy, pandas, gensim), R, MATLAB, Microsoft SQL
- Server, Oracle PLSQL
- Machine Learning Algorithms: classification, regression, clustering, feature engineering
- Big Data Tools: Hadoop, MapReduce, SQOOP, Pig, Hive, NOSQL, Cassandra, Spark
- Others:Deep Learning, NLP, Topic Modeling, Sklearn, Graph Mining, Text Mining, Weka, C, C++ , Java, Javascript, ASP,
- Shell Scripting

Experience

06/2016 to 09/2016
Lockheed Martin Corporation

- Propose a comprehensive neural network model by incorporating user information and underlying event information into document-level Twitter sentiment classification using Deep learning neural networks like CNN, RNN, LSTM, GRU.
- Multiple experiments have been implemented to verify the performance of sentiment classification on a large-scale Twitter dataset using Keras tools on Theano.
- Modeling the March Madness March 2016 - Learn which learning models, data processing technique, ordering technique and parameters tuning yielded the best classification accuracy.
- Using machine learning algorithms like SVM, Naive Bayes, Random Forest, Neural Network, XGboost, etc.

03/2016
Accenture

- Created and explored networks from IMDb movie data - Studied the properties of actor/actress graph and movies graph using PageRank, Community Finding- FastGreedy Newman algorithm, etc., and predicted movie ratings based on regression model and graph model.
- Human Attributes Extraction in Profiles March 2016 - Collect many raw profiles in websites by our configurable crawler.
- Use statistic methods to do pre-processing works include lexical analysis like Tokenization, POS tagging, parsing, and name entity recognition using method of linear chain Conditional Random Field (CRF), and Entities relation extraction.
- Build the semantic patterns, which can be used in model to predict the person attributes.
- Recommendation System using the Movie Lens dataset January 2016 - Predict the rating of all the movies using Collaborative Filtering, alternating least squares and 10 fold cross validation.
- Once the ratings are predicted they can be used to recommend top N movies to each user.
- Popularity Prediction on Twitter January 2016 - Predicted which topics will become popular on Twitter using Linear Regression and Super Bowl hashtags.

03/2012 to 08/2014
Business Analyst Cognizant Technology Solutions ï¼ Yuma , MI

- Supported all business functional components, and provided support assistance to all functional business components.
- Involved in developing Canon Mercury Back-end systems for data transferring between Canon Americas Mercury system in USA and Canon Inc.
- Master Mercury system in Japan using C and Pro.
- C -Assisted in developed Oracle database using Oracle PLSQL of Mercury System, and data migration among Mercury and other Canon IT systems -Produced Unix bash/corn shell scripting to do back-end process operation, system resources checking, and job scheduling.

01/2011 to 02/2012
Business System Analyst Company Name ï¼ Overland Park , State

- Provided web application client support, monitored and reported on client problems and issues.
- Assisted back-end developer for reviewing and debugging C program for network communications in Claims systems and Reproduced Oracle schema objects for Data warehouse.
- Involved in web development of online Claims and Invoices systems using JAVA -Reviewed and reproduced online JAVA reports of Invoices system, checked DB references in it for intelligent Decision Support System -Reproduced schema objects on FindWdev instance of Oracle PLSQL for Data warehouse for historical record and Data mart for reporting purpose, and for Peoplesoft for HR.

06/2010 to 12/2010
Data Scientist/Analyst Company Name ï¼ Novi , State

- Assisted senior quantitative analyst in assessing risk management of exchange rate for FX G-10 products using machine learning techniques for providing appropriate investment recommendations -Mentored sophisticated organizations on large scale customer data and analytics using advanced statistical and machine learning models relying on decision trees, random forest, logistic regression, and other categorical

data, for loan issue.
- Architected and implemented analytics and visualization components for transaction data analysis platform by using d3.js -Performed Unix Shell Scripting for compressing, partitioning and piping large amounts of data.

Education and Training
2016
Master of Science : Computer Science University of California ï¼ City , State , USA Computer Science 3.78 Courses:-Statistics Programming -Databases and Knowledge Bases -Graphs and Network Flows -Language and Thought -Current Topics in Computer Theory -Computer Science Classics -Basic Data Science -Data Mining and Big Data Analytics -System Security
2010
Master of Science : Electrical Engineering Computer Networks -Database Management Systems -Data and Computer Communications -Data Structures University of Bridgeport ï¼ City , State , USA Electrical Engineering 3.62 Computer Networks -Database Management Systems -Data and Computer Communications -Data Structures
2007
Bachelor of Science : Telecommunications Engineering Jilin University ï¼ City , State , China Telecommunications Engineering 3.50
Skills
analyst, ASP, bash, Basic, Big Data, C, C++, clustering, client, client support, data analysis, data processing, data migration, Data Mining, Data Visualization, Data warehouse, Databases, debugging, Decision Support, experiments, functional, Graphs, HR, investment recommendations, Java, Javascript, Machine Learning, machine learning, machine learning, MATLAB, Microsoft Excel, exchange, works, Modeling, NLP, Network, networks, Neural, NOSQL, Oracle, Oracle database, DB, developer, PLSQL, Peoplesoft, POS, Predict, Pro*C, Programming, Python, reporting, assessing risk, scheduling, Shell Scripting, Microsoft SQL Server, Statistics, Tableau, Unix, Unix Shell Scripting, validation, web development, websites