PHAM BA CUONG QUOC

Senior Data Scientist, Kaggle Expert
pbcquoc.github.io
pbcquoc@gmail.com
(+84) 378 050 314

Overview

I'm a Data Scientist, Deep Learning Engineer, Big Data Engineer, Blogger. I am fascinated about deep learning, computer vision. I love to share my knowledge at **pbcquoc.github.io** and contribute to Vietnamese open source.

Work Experience

Data Scientist
Remote Data Scientist at EcoIT, a startup

06.2019 – Now

I lead a team to solve document digitization problems. Our task is a combination of object detection to identify region of interests and character recognition. Our solution is used for government projects with precision more than 95%

Lecture
Lecture at VietAI, Non-profit Organization

10.2018 - 7.2019

I teach Deep Learning for Vietnamese student. My topic is about CNN, LSTM, Attention Mechanism for Machine Translation.

Data Scientist
Data Scientist at Zalo Group, VNG Corporation

09.2017 – 09.2019

I build a large-scale machine learning system that store, process and detect relationships of over 80 million users. The system stores historical data about users location (gps, wi-fi) and their personal information. I extract edge attribute of the super graph (16 billion edges) about their distance in several days, and build a machine learning model using these features to classify their relationship. My model has 86% precision and 0.9 AUC

Our user is annoyed by irrelevant notification of their friend activity. I suggest and build a smart notification system that leverage the relationship and use PageRank to rank tie strength. Our system improves user engagement by 10%

I build a dashboard for our machine learning system that supports AB testing and monitor various measurements. The system helps us reduce experiment time

Data Scientist
Lead Recommender System at PayTV, FPT Telecom

07.2016 – 09.2017

I build a movie recommender system to serve 600k users using PredictionIO, MLLib Spark, Scala. In detail, I implement Alternative Least Square for Collaborative Filtering approach and Single Value Decomposition for Content-Based approach, after that I combine these models using a linear model. My system improves performance to 500% request-per-second and Click-Through-Rate increase 40%.

I proposal and build a music recommender system using Generative Adversarial Networks which is a novel deep learning model. The input to the system is rely on only raw audio so songs can be suggested without any meta-data. The input is converted to MFCC features and pass to my model to learn representation, then i use cosine similarity for ranking songs. My system result is impressive than other based on meta-data

My company business grow, we need a system that must be scalable, flexible, easy to integrate new model, support AB testing which is a well-known business metrics, and lastly, realtime adapt to user when they watch new movies or music. I propose an architecture which combines Spofity's recommender system and Netflix's recommender system, and lead 4 teammates to build the system.
I use user-item logs of our log system to determine linear trendline and suggest to user

I make alert system for our system to track anomaly and email to related person in time.

Education

Bachelor of Science - Honor Program
Ho Chi Minh City University of Science
2012-2016
Thesis advisor: MSc Tran Trung Kien
We implement a novel model to generate image description. Our approach is an association of two most successful deep learning models in Computer Vision and Natural Language Modeling that known as Convolution Network and Long Short-Term Memory. Our result can help blind man to see the world.

Academic Activity

Deep Learning
Ho Chi Minh City University of Science
2015-2016
Our group have learnt more about Supervised Learning based on Bengio's book. It includes mathematics optimization, understanding deep learning, having some experiments to find out deep learning model behavior.

Blog & Github

I try to share my knowledge as much as possible for helping beginners start to learn about AI

Nhận Dạng Chữ Tiếng Việt - Vietnamese OCR	@pbcquoc/vietnamese_ocr
Tìm Hiểu Mô Hình YOLO Cho Bài Toán Object Detection - Understanding YOLO	@pbcquoc/yolo
Tìm Hiểu và Áp Dụng Cơ Chế Attention - Understanding Attention Mechanism	@pbcquoc/attention_tutorial
Tìm Hiểu Convolutional Neural Networks Cho Phân Loại Ảnh	@pbcquoc/cnn
Huấn luyện mô hình CRNN cho nhận dạng chữ viết tay Tiếng Việt - How to train your dragon.	@pbcquoc/crnn

Achievement

2016	The Best Thesis, Ho Chi Minh City University of Science
2017	Top 1 Zalo Hackathon, We build chatbot for selling product
2018	Top 4 Entropy, Data Analytics Competition Top 2 Zalo AI Challenge, The First AI Competition in Viet Nam Top 1 Cinnamon Hackathon, Vietnamese Optical Character Recognition

Skill

Base Knowledge	Deep Learning, Machine Learning Computer Vision, Natural Language Processing Data Analysis, Statistics, Computer Graphics
Programming	Python, Cuda C/C++, MATLAB, Lua, Java, HTML/CSS, C/C++, C#
Libraries	Tensorflow, Theano, Torch7, Scikit-Learn, Pandas
Languages	English (IELTS 6.0)