

Tuan TA

Data Scientist & Principal Engineer

Summary

GeeK and ambitious, contact me for detailed information once you got interesting things!

Experience

Principal Engineer at Deep Signals

May 2016 - December 2017 (1 year 8 months)

Build one of sexiest Fintech applications in the world. We explore insight connected information and predict up coming anomaly / hidden events in large-scale datasets. I am responsible for:

1. Be a part of board members, defining the concept of products and business model.
2. Determining solutions and designing infrastructure for the whole system, which is:
 - Applied many novelty techniques related to different domains, including Machine learning, Deep learning, Graph analytics and Big data.
 - Based on hot technologies on the trend for PoC and production, e.g.: Reactive streaming, Spark, GraphX, Akka toolkit, Hadoop ecosystems, Elasticsearch, Tensorflow, Neo4J and many others...

Lead Data Scientist at Sentifi

November 2014 - May 2016 (1 year 7 months)

Lead a team more than 10 Data scientists / engineers and coordinate tasks with other departments to build core engines for all Sentifi businesses. Solving problems in 4 different areas:

1. Content problems
 - Build Semantic engine based on contextual meaning of financial terms to extract events, topics and impacts inside financial relevance messages.
 - Apply and improve distributed Topic modelling methods take into account automatically defining rules, extracting new events and named / lexicon entities.
 - Detect anomaly topics in different time spans.
 - Build a system which able to automatically optimize forecasting models in real-time to predict hot trending of events in topics.
2. Publisher problems
 - Build classifiers to extract named entities, roles, occupations, and tags of user profiles in different social networks.

- Build a model to extract and/or predicting location information of a profile candidate.
- Build a ranking system, scoring millions qualified profiles in different views.
- Build a model to detect whether a profile is a bot or human.
- Clustering profiles and mapping profiles via crossing social networks.

3. NER related problems

- Extract named entities related to financial worlds via internal data and knowledge bases from communities (i.e. feeds, Freebase, ...)
- Build a classifier to detect whether a message is relevant to finance.
- Build huge number of classifiers for tagging incoming messages to relevant topics.

4. Infrastructure

Design and build Distributed system based on Akka, Play framework, Spark, Elasticsearch and Reactive streaming on top of Hadoop ecosystems. The system is flexible to scale and easy to integrate with multi-tenancy environment. The design follows Service-oriented architecture and Reactive micro-service architecture.

Lead Data Scientist at NEC

July 2014 - January 2015 (7 months)

1. Responsible for leading Big data team (10 to 15 people) in Vietnam, and participating with Data mining team (NEC US Lab) to implement top confidential research projects, which:
 - Developed new analysis techniques, focusing on Mixture model and Deep learning to extract heterogeneous patterns. We achieved fast and high-precision analysis of super-large-scale data.
 - Implemented number of feature transformation & description methods, applying in Feature selection and Feature combination processes.
 - Implemented number of prediction and forecasting methods from among a large volume of candidate factors.
2. In other to improve techniques in agriculture, we developed methods combining predicting and real-time tracking of fishes (includes living environment and growth process) on Android mobile devices.
3. Developed a back-end service which synchronizes large number of photos automatically between servers and camera devices.

Big Data Architect at MobiVi

2013 - July 2014 (1 year 7 months)

1. Reformed legacy system of organization by Distributed system which optimizes data collection activities, concurrent computing and storing for Mobivi backend system.

2. Handled integration processes with Mobivi's partners for e-commerce activities, applying asynchronous I/O frameworks to build web-based applications/services (SOAP & REST).

Researcher at John von Neumann Institute - Vietnam National University HCM

December 2010 - July 2013 (2 years 8 months)

1. Proposed an innovative contextual advertising system driven by image resource on web page, which automatically associates relevant ads with an image rather than the entire text in a web page and seamlessly inserts the ads in the nonintrusive areas within each image. The proposed system represents the attempt towards contextual advertising. The relevant ads are selected based on both textual content and visual content. Our system is able to achieve effective advertising which is a powerful complement to conventional contextual advertising.
2. Developed detect insertion point methods for embedding ad on image for resource-based adaptive advertising. The project funded by VNU-HCMC (code: CS2011-42-02)

Research Intern at National Institute of Informatics

July 2012 - January 2013 (7 months)

- Handled a project which was given by National Institute of Informatics (NII) and National Institute of Genetics (NIG) in order to understand the behavior of zebrafishes by studying their neural activities. I completed 2 phases of the project, which:
- Tracks and extracts spontaneous neuron's activities on biological images (completed in other research by NIG experts), and make analysis on neuronal signals.
 - Implements time-series mining techniques searching the relations between neurons and detect correlations between them (based on actions performed by observed zebrafish).

Android Developer at AntTek

February 2011 - July 2011 (6 months)

Research Assistant

August 2008 - September 2010 (2 years 2 months)

1. Participated with InfoVis research group at Universität Rostock (Germany), completed a number of GeoVis-related projects. These projects focus on implementing general design strategies for visualizing spatially and temporally referenced tags similar to tag clouds on arbitrary-shaped geographical regions. The researches also concentrate on interactive visualization prototypes for the exploration of both spatial and temporal references of Flickr tags.
2. Study, implement and develop a module for the description of image objects concerning image visual features. This project funded by VNU-HCMC (code: B2008-82-05).

Education

University of Science, Vietnam National University - Ho Chi Minh City

Master of Science (MSc), Information and Communications Technology, 2011 - 2013

University of Information Technology, Vietnam National University - Ho Chi Minh City

Bachelor of Science (BSc), Computer Science, 2006 - 2010

Tuan TA

Data Scientist & Principal Engineer



[Contact Tuan on LinkedIn](#)