MARSAN MA （馬源朗）

Phone: (+886) 912-037067
Email: marsan@gmail.com
LinkedIn: https://www.linkedin.com/in/marsanma

Summary

I am a senior engineer with 8-years industry experience, currently working for the top-1 mobile live casting app company "17 media" in Taiwan.

I am always looking for opportunities to leverage machine learning techniques, building predict models and fancy features. My interest including ad-tech, recommender system, image recognition, language models, dialog system.

I have a broad skill set, since I studied electronic engineering, having 4 years experience on system-on-chip, then work on internet startups for another 4 years. I know both software and hardware, technology and business.

Education

Master's degree of Electronic Design Automation
(National Taiwan University, 2005 – 2007)

Bachelor's degree of Electrical and Electronics Engineering
(National Taiwan University, 2002 – 2005)

Honors & Awards

Ranked 45th (top 3.7%) in Facebook Challenge: Predicting Check-Ins
(Jul. 2016, by https://www.kaggle.com/c/facebook-v-predicting-check-ins)

The goal of this competition is to predict which place a person would like to check into. For this competition, Facebook released 40M check-in data of a small city, and our task is to predict the most likely check-in places of 8M users.

Since the 40M samples is huge, it also challenging us about how to deal with large scale data. I've used tricks on dividing the question size without losing too much predicting accuracy, and speed-up, parallelize every code detail.

My final model is a large blending model composed of stacked random forest and probability model, K-nearest neighborhood, and gradient boost trees. This model achieved map@3 = 0.59856 (while the 1st prize winner is 0.62279)

Ranked 30th (top 10%) "Truly Native?" Kaggle challenge
(Oct. 2015, by https://www.kaggle.com/c/dato-native)

In this challenge, we are requested to develop a system predicting whether an article is a native advertisement. It's involves lots of natural language processing tricks, like boilerplate removing, topic modeling and embedding techniques. My final model is a blending of several bagging logistic regression, two gradient boost trees, and two factorized machine, which achieved 0.974 in AUC (while 1st prize winner achieves 0.998).

"Best Tech Award" in IDEAS TECH Hackathon 2014
(Jul. 2014, by Institute for Information Industry, Taiwan)

In this hackathon, I build a forum topic analysis website http://piposay.com which gather the news from mainstream medias and rank them in popularity among social networks, and cluster articles into topics using topic modeling algorithms. Furthermore, I programmatically summarize articles into excerpts around 150 words, extract celebrities to be his/her biography through named-entity-recognition techniques.

"First Prize" in Computer Aided Design of Integrated Circuit contest
(Jul. 2006, by Ministry of Education, Taiwan)

In this competition, we are requested to solve a problem in integrated circuit manufacturing called “latch-up check”. There are two kinds of instances on IC-manufacturing where if they are too close to each other in the layout, they tend to short. So we need to check whether they are far enough from each other. We apply the Dijkstra's algorithm with a little modification. And the final results we pass all the test cases with the shortest computing time.

Experience

Data Scientist of 17 Media
(Apr. 2016 – Present)

As the first data scientist, I plan and implement the data flow and infrastructure for the inner analyzing platform, conclude raw data all the way into the real-time dashboard and insights for operation and marketing team, and applied my good old machine-learning works on content analysis with DNN and NLP techniques.

I also built the recommendation engine which out-performed the vanilla ALS algorithm used by Spark, achieve MAP@3 around 0.09, which is far better than the benchmark (recommending most popular items, MAP@3 around 0.03).

Data Scientist of cacaFly/TenMax AD Tech Lab
(Oct. 2014 – Mar. 2016)

cacaFly is the Top1 digital advertisement company in Taiwan. Here I build infrastructure for machine learning and data mining, and integrate the latest machine learning results into our ad-tech products, like bid-price prediction, advertisement performance prediction and visualizing history trend. I also mentor our data scientist interns from NCTU.

TenMax AD Tech Lab is funded by cacaFly, focus on RTB (real-time-bidding) product. Since in RTB we have the features of everyone single impression (user viewing a page), we could do detail profiling on both user and web context, through applying state-of-the-art algorithms like FTRL-P, natural language processing and collaborative topic-modeling and deep-neural-network.

Founder of PipoSay (http://piposay.com)
(Jan. 2014 – Oct. 2014)

I build this news aggregation service focus on trend analysis, ranking hot topics and gather latest stories of your favorite people. No human editors, all content and data are programmatically processed in machine learning and natural language processing techniques. This work wins the 2014 IDEAS Tech Hackathon "Best Tech Award" held by Taiwan government.

Technical Director of Condé Nast Taiwan Wired (http://wired.tw)
(Nov. 2012 - Nov. 2013)

Our plan was to build user profiling engine from medias since only medias themselves could control what they are going to publish. If they know exactly what their audiences want, they could publish with content prone to be popular.

So I built a platform which analyze the popularity of new articles all over the world, and show what's the real-time hottest topic, and what's the booming topic that only a few people have seen, but with a lot of comments and share counts on social networks. This inner information saves a lot of time for our editors since they could easily choose the topic to write, and the topic is almost guaranteed to be popular among our audiences.

We pushed the Alexa rank of WIRED-Taiwan from #1100 to #500 in 6 months.

Senior Engineer of Novatek (http://www.novatek.com.tw/)
(Jul. 2008 - Feb. 2012)

NovaTek is the 2nd largest SoC design house and the largest driver IC vendor in Taiwan. I had been in charge of video codec IP design, SoC design & integration. Video/Image related algorithm development and circuit implementation.

Been a system-on-chip designer for 4 years, I am familiar with the advantages and weaknesses of various hardware modules and devices. This makes me a better engineer since I actually know what really happened in hardware behind the software code, how to speed-up and how to avoid traps while coding software.

Certifications

Certificate Japanese Language Proficiency (Feb. 2007)

Japanese Interchange Association, License 2X025045

Certificate of PADI Open Water Diver (Jul. 2009)

Professional Association of Diving Instructors, License OWSI-489829

Patterns

Shutter glasses, three-dimensional imaging system and method for controlling shutter glasses (Taiwan NVA-P0296-CHN:0) (Feb. 2011)

A shutter glasses for a three-dimensional video system includes a first liquid crystal lens, a second liquid crystal lens, a lens control module, coupled to the first liquid crystal lens and the second liquid crystal lens, for alternatively blocking the first liquid crystal lens and the second liquid crystal lens according to a synchronization signal, an image capturing device, for capturing a plurality of images including a synchronization image outputted from a liquid crystal display of the three-dimensional video system, and a timing generating module, coupled to the image capturing device, for generating the synchronization signal according to the synchronization image.

Languages

Chinese (Native)

English (Proficient, TOEIC 910)

Japanese (Workable, N2 level)

Programming Skills

Machine Learning (main skill)

TensorFlow (experienced)

Natural Language Processing (experienced)

Python (proficient, main language)

Ruby on Rails (proficient)

AWS / Linode (depend on the budget)

Docker (infra of must)

Spark (promising but expensive, willing to try SMACK)

R, C, Java, php (available on request)