Zijian Zhang

Curriculum Vitae

Essentially, all models are wrong, but some are useful.
–George E. P. Box

Education

2015.03–2018.03 **Masters of Science**, *Leibniz Universität Hannover*, Hannover, Lower Saxony, Germany, *weighted average examination notes: 1.79. The score including master thesis will be 1.56~1.76 presumably.*
Informatik (Computer Science)

2014.09–2015.03 **Language Courses (German)**, *in various of language schools*, Hannover, Lower Saxony, Germany, *DSH-3 (Deutsche Sprachprüfung für den Hochschuluzugang - Stufe 3)*.
German language

2010.09–2014.06 **Bachelor of Science**, *Xidian University*, Xi'an, Shaanxi, China.
Information Security

2007.09–2010.06 **Senior High School**, *Tangshan 1st Senior High School*, Tangshan, Hebei, China.

2004.09–2007.06 **Junior High School**, *Tangshan 12th Junior High School*, Tangshan, Hebei, China.

1998.09–2004.06 **Elementary School**, *Lubei District Experimental Elementary School*, Tangshan, Hebei, China.

Birthday

23, May, 1992.

Family Status

Single.

Master Thesis

Title **Scalable Approaches for Learning Word Representations**

Supervisors Prof. Dr. Avishek Anand (anand@l3s.de) &
Prof. Dr. techn. Dipl.-Ing. Wolfgang Nejdl (nejdl@kbs.uni-hannover.de)

Mendelssohn Str. 26c – 30173 Hannover, Lower Saxony
(0049) 157 3344 6323 • ✉ zhangzijian0523@gmail.com
Einfaches Chalet (Chinese version only currently)
https://github.com/JoshuaGhost

Description Propose an approach of training distributed word representation model in a scalable way, which significantly accelerates the training process of a word2vec model on huge amount of corpora by parallelization, meanwhile its performance is mostly preserved.

Bachelor Thesis

Title **GPU Accelerated Detection of Communities in Graph**

Supervisors Prof. Qi, Yutao (ytqi@xidian.edu.cn) &
Prof. Zhang, Weidong (wdzhang@xidian.edu.cn)

Description This thesis propose a CUDA based approach of accelerating communities detection and division in graph data. This approach greatly makes use of the ability of high performance parallel computation of a GPU and diminish the duration of community detection in tens of times

Experience

Miscellaneous

2016.08.01–2018.04.30 **Student assistant job (HiWi) with Master Thesis**, Forschungszentrum L3S, Hannover, Lower Saxony, Germany.

Propose an approach of scaling training process over gigantic corpus. The original SGNS Word2Vec model is thoroughly inspected and a strategy of divide-and-combine method of training is figured out for dealing with large corpus. In the beginning the original corpus is divided into several word/bigram-distribution-preserving sub-corpora and then a Word2Vec model are trained based on each individual sub-corpus independently. At last several combination method for sub-models have been conducted and evaluated. For those model combination method which produces a high-dimensional model, a following PCA for dimension reduction is attached at the end of framework. This project became later on the master thesis of mine. Furthermore, in order to tackle some interesting problems in the future work, for instance unbalanced vocabulary problem, and make notable impact for the territory, I decided to continue contributing to this project even after the master thesis is handled over.

Detailed achievements and gains:

- Master programming with Python in a pythonic way;
- Learned several machine learning algorithms, including but not limited to:
  - Classification algorithms, from basic Naïve Bayes classifier, C45 classifier and SVM etc. to more advanced kernel based models,
  - Regression strategies like MSE and logistic,
  - Clustering algorithm such as K-means, hierarchical agglomerative/divisive clustering, density based clustering etc.,
  - Generative models, especially in the aspect of word embedding such as GloVe, word2vec etc.,
  - Foundation of optimization theory,
  - Foundation of school of Bayes, including but not restricted to MCMC, Metropolis-Hastings, ME algorithms, together with their mathematical foundation including Lebesgue-Stieltjes integral in real analysis, fundamental thought of variational in functional analysis, Markov model and LDE in stochastic process.
  - Classical algorithms in linear algebra (mainly dimension reduction algorithms) such as PCA, MDS, Laplacian matrix and spectrum analysis of graph etc.,
  - Algorithms such as (General) Canonical Component Analysis/Canonical Variance Analysis, (General) Procrustes Problem, Tensor Canonical Component Analysis, which contribute to integrate different view of same dataset.
  - Knowledge on scientific English writing;
- Foundation of deep neural network, including but not limited to:
  - Classical models like Restricted Boltzmann Machine, Quantum Boltzmann Machine, auto encoder, deep belief network etc.;
- Basic scalable programming on Hadoop cluster and MapReduce Streaming, operating Hadoop HDFS etc.;
- Foundation of information retrieval;
- Foundation of NLP, especially in realm of Word-Embedding.

2015.09.01–2016.07.31 **Studentische Hilftskraft**, Institut Mess- und Regelung (Institute of measuring and controlling), Leibniz Universität Hannover, Hannover, Lower Saxony, Germany.

Contributed to a doctor project about automatic detecting defection on surface of manufactures using laser diode. Help to build a prototype of project using C/C++, Visual C++ and Matlab.

Detailed achievements:

- Learned OS relevant C/C++ programming e.g. inter-processes communication etc.;
- Learned basic Matlab science numerical calculation programming;
- Learned GUI designing within Matlab using including GUIDE and Matlab GUI framework.

2016 **Project on Agile developing**, Institut Softwareentwicklung (Institute of Software Development), Leibniz Universität Hannover, Hannover, Lower Saxony, Germany.

A half-year-long software projection in the institute of software development, during which ability of cooperation with a German-speaking-scrum-team as well as terminologies and practical implementation of agile software development is trained

Detailed achievements:

- Knowledges on agile development and methods within, such as scrum, Kanban method, paired programming etc.
- Knowledges on Java programming language.
- Ability of independent working and acting as team player.

2013–2014 **Student Part-time job**, School of Electronic Engineering, Xi'an, Shaanxi, China.

Analyzing functionality of ether/wireless net communication protocols, designed the in order to enhance transmission security during communication as well as camouflage high-valuable-targets.

- Work-flow of 802.x series of protocol and communication encryption within
- Working principle of Intranet sniffer, including the host discovery process and protocol fingerprint technique etc. and strategies of camouflage the high-valuable-hosts in front of sniffers.
- Knowledge of GNU/Linux kernel programming, including kernel modules, KDB, netfilter/IP tables etc.

Awards

2012 National 2nd Prize in The National Undergraduate Electronic Design Contest Information Security Invitational 2012

2011 National 2nd scholarship in Xidian University

2007–2009 2nd Prize in NOIP (National Olympiad in informatics in Provinces) for three times

Computer skills

Programming	C/C++, Python, Java, Pascal, Matlab, GNU Bash
OS	GNU/Linux, including daily management and kernel programming
Theories	Algorithms in graph, Dynamic Programming, Fundamental machine learning and neural networks, Basic knowledge on quantum mechanic
Front-end	Matlab GUI Toolkit, QT

Mendelssohn Str. 26c – 30173 Hannover, Lower Saxony
(0049) 157 3344 6323 • ✉ zhangzijian0523@gmail.com
Einfaches Chalet (Chinese version only currently)
https://github.com/JoshuaGhost