Young-Jun Ko

youngjun.ko@gmail.com | +41 78 63 33 55 8 | yjk21.github.io

Address: Balberstrasse 68, 8038 Zurich, Switzerland
Citizenship: German
Work Permit: Swiss Permit C

Current Occupation

Since Nov.2018 **AI DevTech Engineer at NVIDIA - Performance Software Engineering for Deep Learning Workloads**

Optimized CUDA kernels for MLPerf BERT Training v1.0 (Link)

Optimized CUDA kernels for MLPerf BERT Inference v0.7 (Link)

TensorRT-based accelerated inference of BERT (NVIDIA Dev Blog Link)

Contributed GLM CUDA code to RAPIDS cuml (Link)

Education

Feb.2011 - Feb.2017 **PhD at EPFL, Switzerland**

Supervisors: Prof. Matthias Grossglauser, Dr. Matthias W. Seeger

Thesis: Applications of Approximate Learning and Inference for Probabilistic Models

Oct.2007 - Jun.2010 **MSc in Comp. Sci. at Saarland University, Germany (GPA[1]: 1.3)**

Supervisor: Prof. Matthias Hein

Thesis: Stability of Feature Selection Methods

Oct.2002 - Sep.2005 **Dipl.-Inf.(BA)/BSc(OU) in Applied Computer Science at the University of Cooperative Education, Germany (GPA: 1.4)**

Thesis: Rule-based Invoicing of IBM Infrastructure Services for Deutsche Bank

1993 - 2002 **Abitur at Lessing Gymnasium, Germany (GPA: 1.3)**

Publications

2016 *YJ Ko*, L. Maystre, M. Grossglauser, Collaborative Recurrent Neural Networks for Dynamic Recommender Systems, *ACML 2016*

2015 *YJ Ko*, M. Seeger, Expectation Propagation for Rectified Linear Poisson Regression, *ACML 2015*

2014 *YJ Ko*, M. E. Khan, Variational Gaussian Inference for Bilinear Models of Count Data, *ACML 2014*

M. E. Khan, *YJ Ko*, M. Seeger, Scalable Collaborative Bayesian Preference Learning, *AISTATS 2014*

2013 M. E. Khan, *YJ Ko*, M. Seeger, Scalable Bayesian Preference Learning, *NIPS 2013* Workshop on Personalization 2013

2012 *YJ Ko*, M. Seeger, Large Scale Variational Bayesian Inference for Structured Scale Mixture Models, *ICML 2012*

[1] All GPAs on a German scale: 1.0 (Best) - 6.0

Scholarships, other Professional Activities

NVIDIA NTECH 2020 Presented work on optimized Multi-head Attention

CSCS Summer School 2020 Lecturer

CSCS Summer School 2019 Lecturer

ACML 2017	Senior PC member
Apr.2015	Google internship offer, Mountain View, California
Jul.2010 - Sep.2010	PhD Scholarship of the International Max Planck Research School for Computer Science (Cancelled due to the move to EPFL)
Apr.2008 - Sep.2009	Scholarship of the Saarbrücken Graduate School in Computer Science
Summer 2008	Tuition Fee Waiver (Grades in the top 5%, afterwards fees were abolished)

Work Experience

Apr.2017 - Sep.2018	**Senior Machine Learning Engineer at 1plusX AG, Zurich, Switzerland**
	Representation learning and predictive modeling based on interaction events using neural networks in Scala/Spark and Python
Feb.2011 - Feb.2017	**Research- and Teaching Assistant at EPFL**
Jun.2010 - Feb.2011	**Research Assistant at Saarland University**
Oct.2006 - Oct.2007	**Software Developer at SAP AG - Analytical Banking**
	SAP Bank Analyzer 6 Process Infrastructure Development in ABAP Objects on the Netweaver Plattform
Oct.2005 - Sep.2006	**Software Developer at IBM Systems & Technology Group**
	IBM Total Storage Productivity Center Development (Storage Area Network Management based on SNIA SMI-S Interface)
Jan.2005 - Mar.2005	**Internship at IBM ITS Consulting & Architecture**
	Development of a network connection analysis tool in J2EE
Jun.2004 - Sep.2004	**Internship at IBM AMS Solution Implementation**
	Design and implementation of a framework for backend system integration
Jan.2004 - Mar.2004	**Internship at IBM BCS Sector Finance**
	Design and implementation of a web framework for frontoffice work places

Languages and Technologies

Languages:	German (native), English (fluent), Korean (intermediate), French (basic), Italian (basic), Latinum
Programming:	Python, Julia, Matlab, Scala, C/C++, Java, SQL
Technologies:	MxNet/Gluon, Theano and Tensorflow, GPU computing (CUDA), Apache Spark, Numpy/Scipy, Pandas, Linux, git, web (flask, javascript), Apache Kafka

Zurich, May 15, 2021