Yongsub Lim Founder & Principal Data Scientist at MakinaRocks

Contact

- **Email**: *yongsub (at) makinarocks.ai*
- **US**: +1–408–747–9493 ｜ 310 De Guigne Dr, Sunnyvale, CA 94085
- **KR**: +82–10–9164–2469 ｜ 6F, 311 Gangnam-daero, Seocho-gu, Seoul 06628

Education

- **Doctor of Philosophy**, Computer Science, KAIST   Aug 2015
  Research Topics: graph mining, data mining and stream mining.
  Advisor: U Kang

- **Master of Science**, Computer Science, KAIST   Feb 2011
  Research Topics: inference in graphical model and social network analysis.
  Advisor: Kyomin Jung

- **Bachelor of Science**, Information and Computer Engineering, Ajou University   Feb 2009

Professional Experiences

- **Founder & Principal Data Scientist**   May 2018 ∼ Present
  MakinaRocks

  As a principal data scientist, I focus on conducting data analytics projects and developing software assets such as a core AI engine of MakinaRocks.

  - I carried out a project to predict failures of equipment used in semiconductor manufacturing. Based on our anomaly detection model, our team devised a new technique to estimate the remaining time to failure from the time when anomaly is detected.
  - I led a research project related to anomaly detection based on deep neural networks. This work proposed a new metric to measure abnormality, which fully utilizes information discovered by deep neural networks.
  - I'm responsible of developing a part of our core AI engine, which is related to automation of data analytics.

- **Data Scientist** @ Data Analytics CoE   Feb 2017 ∼ May 2018
  Data R&D Center, ICT R&D Center, SK Telecom

  As a member of PdM (Predictive Maintenance) Group, my focus was on applying machine learning techniques to industrial data to solve problems arising in real environment.

  - I carried out two projects for anomaly detection in semiconductor manufacturing data. In the first project, our team evaluated various models in various aspects in terms of anomaly detection, and in the second project, developed a deep learning based model which results in detecting most anomalies in data.
  - I also participated in a research project to develop a method to maximize model performance for anomaly detection when only few abnormal cases are given, resulting in publishing a paper.

- **Postdoctoral Researcher** @ Data Mining Lab   Sep 2015 ∼ Jun 2016
  Department of Computer Science and Engineering, Seoul National University

Awards & Honors

1. **Winner** of the grant award   Jan 2014
   Venture Research Program for Graduate and PhD Students from KAIST

2. **Honor Prize**   Feb 2013
   The 19th Samsung Electronics Humantech Thesis Prize from Samsung

Journals

1. Minsoo Jung, **Yongsub Lim**, Sumin Lee, and U Kang. *FURL: Fixed-memory and Uncertainty Reducing Local Triangle Counting for Multigraph Streams*. Data Mining and Knowledge Discovery (DMKD), 2019 (to appear).

2. **Yongsub Lim**, Minsoo Jung, and U Kang. *Memory-efficient and accurate sampling for counting local triangles in graph streams: from simple to multigraphs*. ACM Transactions on Knowledge Discovery from Data (TKDD), 2018.

3. **Yongsub Lim**, and U Kang. *Time-weighted counting for recently frequent pattern mining in data streams*. Knowledge and Information Systems (KAIS), 2017.

4. Woosang Lim, Jungsoo Lee, **Yongsub Lim**, Doo-Hwan Bae, Haesun Park, Dae-Shik Kim, and Kyomin Jung. *Hierarchical ordering with partial pairwise hierarchical relationships on the macaque brain data sets*. PLoS ONE, 2017.

5. Heeyoung Kwak, Joonyoung Kim, **Yongsub Lim**, Shin-Kap Han, and Kyomin Jung. *Centrality fairness: Measuring and analyzing structural inequality of online social network*. Journal of Internet Technology (JIT), 2017.

6. **Yongsub Lim**, Won-Jo Lee, Ho-Jin Choi, and U Kang. *MTP: Discovering high quality partitions in real world graphs*. World Wide Web Journal, 2016.

7. **Yongsub Lim**, U Kang, and Christos Faloutsos. *SlashBurn: Graph compression and mining beyond caveman communities*. IEEE Transactions on Knowledge and Data Engineering (TKDE), 2014.

8. **Yongsub Lim**, Kyomin Jung, and Pushmeet Kohli. *Efficient energy minimization for enforcing label statistics*. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2014.

Conferences

1. **Yongsub Lim**, and U Kang. *MASCOT: Memory-efficient and accurate sampling for counting local triangles in graph streams*. 21st ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2015.

2. Adrian Kim, Kyomin Jung, **Yongsub Lim**, Daniel Tarlow, and Pushmeet Kohli. *Minimizing expected losses in perturbation models with multidimensional parametric min-cuts*. 31st Conference on Uncertainty in Artificial Intelligence (UAI), 2015.

3. **Yongsub Lim**[*], Won-Jo Lee[*], Ho-Jin Choi, and U Kang. *Discovering large subsets with high quality partitions in real world graphs*. Second International Conference on Big Data and Smart Computing (BigComp), 2015.   ([*] These authors contributed equally to this work)

4. **Yongsub Lim**, Jihoon Choi, and U Kang. *Fast, accurate, and space-efficient tracking of time-weighted frequent items from data streams*. 23rd ACM International Conference on Information and Knowledge Management (CIKM), 2014.

5. **Yongsub Lim**, and Kyomin Jung. *Decentralized control for intelligent robot system to avoid moving obstacles*. Fourth International Conference on Intelligent Systems, Modelling and Simulation (ISMS), 2013.

6. **Yongsub Lim**, Kyomin Jung, and Pushmeet Kohli. *Energy minimization under constraints on label counts*. The 11th European Conference on Computer Vision (ECCV), 2010.

Workshops

1. Andre S. Yoon, Taehoon Lee, **Yongsub Lim**, Deokwoo Jung, Philgyun Kang, Dongwon Kim, Keuntae Park, and Yongjin Choi. *Semi-supervised learning with deep generative models for asset failure prediction*. 2nd ACM SIGKDD Workshop on Machine Learning for Prognostics and Health Management (ML for PHM), 2017.

2. **Yongsub Lim**, Kyomin Jung, and Pushmeet Kohli. *Constrained discrete optimization via dual space search*. NIPS Workshop on Discrete Optimization in Machine Learning 2011: Uncertainty, Generalization and Feedback (DISCML), 2011.

PATENTS

1. U Kang, Minsoo Jung, Sunmin Lee, **Yongsub Lim**. *Triangles Counting Method for Graph Stream*. Korean patent number: 10-2016-0183070 (filed on Dec. 29, 2016).

2. Dong-Min Seo, Seok-Jong YU, Min-Ho Lee, U Kang, **Yong-Sub Lim**, In-Jae Yu, and Sael Lee. *Method and Apparatus for Network Clustering*. Korean patent number: 10-2016-0101611 (filed on Aug. 10, 2016).

3. U Kang, **Yongsub Lim**, Christos Faloutsos. *Method and Apparatus for Processing Graph Compression*. Korean patent number: 10-2014-0052401 (filed on Jul 28, 2014).

TALKS

1. *Stream data mining: algorithms and applications*. The 42nd Korea Software Congress (tutorial at one of the largest domestic conferences), Pyeongchang-gun, Gangwon-do, Korea, Dec 2015.

2. *Constrained MAP inference algorithm with bi-dimensional parametric minimum cuts*. The 14th Korea-Japan Workshop on Algorithms and Computation (WAAC), Busan, Korea, Jul 2011.

3. *Submodular minimization with constraints using Lagrangian multipliers*. Research on Software Analysis for Error-free Computing center workshop, Paju, Korea, Jun 2011.

4. *Energy minimization under constraints on label counts*. Research on Software Analysis for Error-free Computing center workshop, Sokcho, Korea, Aug 2010.

5. *Graph cut algorithm with application to computer vision*. Winter School on Algorithms and Combinatorics, Yeongi, Korea, Jan 2010.

PROFESSIONAL SERVICES

Journal & Conference Reviews

1. ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2016.
2. International World Wide Web Conference (WWW), 2016.
3. International Conference on Big Data and Smart Computing (BigComp), 2016.
4. ACM/SIGAPP Symposium on Applied Computing (SAC), 2016.
5. ACM International Conference on Web Search and Data Mining (WSDM), 2016.
6. World Wide Web: Internet and Web Information Systems, 2015.
7. IEEE International Conference on Big Data (IEEE BigData), 2015.
8. IEEE International Conference on Data Mining (ICDM), 2014–2015.
9. ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2015.
10. IEEE International Conference on Data Engineering (ICDE), 2015.
11. International Conference on Extending Database Technology (EDBT), 2015.
12. European Conference on Machine Learning and Principles of Knowledge Discovery in Databases (ECML/PKDD), 2015.

TEACHING EXPERIENCES

Guest Lectures

- Memory-efficient Real-time Data Stream Mining: Discovering Recently Frequent Items and Estimating Local Triangles
  - Advanced Data Mining (KAIST)   Jun 5, 2015
  - Machine Learning and Data Mining (KAIST)   Jun 1, 2015

- Graph Compression for Real World Graphs - SlashBurn
  - Advanced Data Mining (KAIST)   Mar 11, 2015
  - Data Mining (KAIST)   Dec 10, 2014
  - Introduction to Web Science Technology (KAIST)   Apr 17, 2014
  - Advanced Data Mining (KAIST)   Mar 12, 2014

Teaching Assistants

- Algorithms: Design and Analysis (KAIST)   Spring 2010 $\sim$ 2013
- Mathematics for Information Science (KAIST)   Spring 2012
- Network of Things (KAIST)   Fall 2011

REFERENCES

Prof. U Kang
Department of Computer Science and Engineering
Seoul National University
Seoul, Korea
ukang@snu.ac.kr

Prof. Kyomin Jung
Department of Electrical and Computer Engineering
Seoul National University
Seoul, Korea
kjung@snu.ac.kr

Dr. Pushmeet Kohli
Machine Learning and Perception Group
Microsoft Research
Cambridge, UK
pkohli@microsoft.com