

Kuan-Liang Liu, Ph.D.

Linkedin profile: [tw.linkedin.com/in/kuanliangliu](https://www.linkedin.com/in/kuanliangliu)
No.1 Lane 31 Kewu Rd. Sanmin district Kaohsiung City Taiwan
Email: kliangliu@gmail.com / Mobile: +886-939320272

SUMMARY

A recent Data Mining graduate with Information Management PhD degree from Taiwan National Cheng Kung University, and great deal of experiences in large scale data analysis, ranging from NGS genomics data (shotgun metagenomic binning, amplicon taxonomic assignment to Human Microbiome Project), social network analysis (Twitter, Facebook influencer marketing supervised learning) to image/text classification (deep learning); seeking a data scientist position to further develop expertise and experience and to solve real world problems.

EDUCATION

Ph.D. in Information Management / National Cheng Kung University *Sep. 2007 ~ Feb. 2015*

- Advisor: Dr. Tzu-Tsung Wong
- Dissertation: Naïve Bayesian classifiers with multinomial models and noninformative generalized Dirichlet priors for rRNA taxonomy assignment

M.S. in Information Management / National Cheng Kung University *Sep. 2005 ~ Jun. 2007*

- Advisor: Dr. Tzu-Tsung Wong
- Dissertation: A Probabilistic mechanism based on clustering analysis and distance measure for subset gene selection

B.S. in Industrial and Information Management / NCKU *Sep. 2001 ~ Jun. 2005*

PROFESSIONAL EXPERIENCE

Business Consultant - AP Region *July, 2012 ~ present*

Viking Products, Inc.

- Supplier Management and Quality Audit
- Problem Solving at supplier site
- QMS Audit
- Supplier QMS Improvement
- APQP, PPAP, MSA, SPC, FMEA assistance for suppliers

Research Assistant *July, 2012 ~ present*

Inst. of Information Management, National Chen Kung University, Taiwan

- Project: Generalized Dirichlet priors for naïve Bayesian classifiers with multinomial models
 - Improve the accuracy of naïve Bayesian classifier for taxonomic assignment by incorporating multinomial model and Generalized Dirichlet priors. 2 rRNA sequence sets, bacterial 16S and fungal 28S with different read lengths (200, 400 and 800bp to mimic sequence read generated by different NGS platform, e.g., Illumina HiSeq and MySeq) were used for experimental test. <http://sourceforge.net/p/gdrnaclassifier/wiki/Home/>

Visiting Scholar

July. 2011 ~ Sep. 2011

Bioinformatics team, Tokyo Tech University, Japan

- Mentor: Dr. Ken Kurokawa and Dr. Hiroshi Mori
- Project: Visualization of 16S rRNA taxonomy assignment
 - Build on VITCOMIC (Visualization tool for Taxonomic COMpositions of Microbial Community, <http://mg.bio.titech.ac.jp/vitcomic/>) and develop a visualization tool compare 16S rRNA taxonomic assignment results (by primers and read length) with hierarchy relationship.

Visiting Scholar

Jan. 2011 ~ May. 2011

B6 (Bioscience group) Los Alamos National Laboratory, USA

- Mentor: Dr. Gary Xie, Dr. Patrick Chain and Dr. Cheryl Kuske
- Project: REVT - rRNA Experimental Validation Test/Tool
 - A web application implemented by Django (python) framework for validating rRNA taxonomic assignment results generated by different taxonomic assignment parameters, including primer region and read length.
- Project: Fungal LSU and ITS sequence database and classifier
 - Compare the accuracy of taxonomic assignments at the Genus, Family, Order and Phylum level using a leave-one-out cross validation (LOOCV) strategy, and compared the use of a computationally fast naïve Bayesian classifier with a computationally intensive BLAST approach. The publicly available database was resourced to multiple web-based sequence analysis sites (e.g. Ribosomal Database Project, Mothur, Qiime) that are commonly used for environmental sequence analysis. <https://rdp.cme.msu.edu/classifier/classifier.jsp>

Short term research & Graduate Research Assistant

Oct. 2008 ~ Sep. 2010

Information Science group at Los Alamos National Laboratory, USA

- Mentor: Dr. Nickolas Hengartner & Dr. Gary Xie
- Project: Metagenomic binning (classification & clustering)
 - Taxonomic assignment, OTU analysis, microbial community comparison on human and environmental microbiome data sets.
 - Utilize markov model to remedy conditional independent assumption on naïve Bayesian classifier for rRNA read taxonomic assignment.

Research Assistant

Sep. 2007 ~ Apr. 2008

Inst. of Biodiversity, National Chen Kung University

- Advisor: Dr. Huan-Lin Wu
 - Administered Unix/Linux machine
 - Developed algorithms with Perl for sequence function prediction

Research Assistant

Sep. 2006 ~ Jun.2007

Inst. of Information Management, National Chen Kung University

- Advisor: Dr. Tzu-Tsung, Wong
- Project: A Probabilistic Mechanism Based on Clustering Analysis and Distance Measure For Subset Gene Selection
 - Developed gene selection algorithm, including DBSCAN clustering algorithm implementation with Visual Basic

SKILLS

- Programming in languages including Python, R, Perl, Ruby and Visual Basic
- Big Data processing: MapReduce, Hadoop, Pig, Hbase, Apache Spark
- Web applications: HTML, CSS, Flot, Django, Ruby on Rails
- Data mining, Machine Learning algorithms and statistical analysis
 - Supervised Learning: Naïve Bayesian Classifier, KNN, SVM, decision tree and applications, Neural Network
 - Unsupervised Learning: Deep Learning, K-means, DBSCAN, OTU analysis
 - Multivariate statistical analysis
- Process Mining – Process discovery, process modeling, bottleneck analysis, conformance checking, social network analysis with ProM and Disco.
- NGS data analysis (quality filtering, chimera checking, OTU assignment, taxonomic assignment, PCA):
 - Pipeline usage: QIIME, Mothur
 - Taxonomic profiling: MEGAN, MetaPhlAn, RDP classifier
 - Functional profiling: ShortBRED, HUMAnA
- Database: SQL, NoSQL
- Quality Management System
 - System certification: ISO 9001:2008 (RABQSA lead auditor), ISO/TS 16949
 - Core Tools: SPC, MSA, FMEA, APQP, PPAP
 - VDA 6.3, CQI-9
- Operation system: Windows, Linux and MacOS
- Badminton for 14 years

AWARDS, HONORS, SCHOLARSHIP and TRAINING

- Academic
 - 2014 - Acceptance for Strategies and Techniques for Analyzing Microbial Population Structures (STAMPS) workshop at Marine Biology Laboratory (MBL) USA.
 - 2011 - Acceptance and Scholarship for Japan Summer Program at Graduate School of Bioscience and Biotechnology, TokyoTech University
 - 2011 - The winner of ISME (International Symposium of Microbial Ecology) travel grant award
 - 2011 - The winner of HMRC (Human Microbiome Research Conference) travel grant
 - 2010 - Acceptance and Scholarship for PSLC (Pittsburgh Science of Learning Center) summer school at Carnegie Mellon University in Pittsburgh
 - 2010 – Acceptance and travel grant for 2010 GMOD (Generic Model Organism Database) summer school.
 - 2010 - Scholarship of Transnational Study & Research Scholarship sponsored by National Cheng Kung University. CR98087
 - 2009 - Scholarship of Graduate Student Study Abroad Program sponsored by National Science Council. NSC97-2917-I-006-111
 - 2008 - Scholarship of Transnational Study & Research Scholarship sponsored by National Cheng Kung University. SR97018
 - 2007 - Scholarship of 1st-place award for Ph.D. entrance exam.
 - 2007 - Scholarship of Excellent graduate student, Institute of Information Management, National Chen Kung University, Taiwan
 - 2006 - Master Thesis Scholarship of Archilife Research Foundation
 - 2006 - Honorable Mentioning List of the SME e-Commerce Model Contest
 - 2006 - Second prize of the 3rd Green Supply Chain Training Course
- Notable fencing results
 - 2009 - Gold medal at New Mexico Badminton double & single state games, ABQ NM
 - 2009 – Gold medal at Colorado Badminton double state game
 - 2004 - Final eight of And1 Basketball games, Tainan Taiwan
 - 2001 - The fourth place of NCKU folk dance competition

PUBLICATION

- Journal Paper

T. T. Wong, **K. L. Liu**, “A Probabilistic Mechanism Based on Clustering Analysis and Distance Measure for Subset Gene Selection”, Expert System With Application, 2010, 37(3), pp. 2144-2149.

G. Xie, P.S.G. Chain, C.-C. Lo, **K. L. Liu**, J. Gans, J. Merritt, F. Qi, “Community and gene composition of a human dental plaque microbiota obtained by metagenomic sequencing”, Molecular

Oral Microbiology, 2010, 25(6), 391-405.

F. Yang, X. Zeng, K. Ning, **K. L. Liu**, C. C. Lo, W. Wang, J. Chen, D. Wang, R. Huang, X. Chang, P. S. Chain, G. Xie, J. Ling and J. Xu, “Saliva microbiomes distinguish caries-active from healthy human populations”, *The ISME Journal*, 2012, 6, 1-10

K. L. Liu, C. R. Kuske, A. P. Alfaro, S. Eichorst and G. Xie, “Accurate, rapid classification of fungi using the large subunit rRNA gene”, *Applied and Environmental Microbiology*, March 2012, 78(5) 1523-33.

K. L. Liu, T. T. Wong, “Naïve Bayesian Classifiers with Multinomial Models for rRNA Taxonomic Assignment”, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2013, 10(5), 1334-9.

A. P. Alfaro, **K. L. Liu**, C. R. Kuske and G. Xie, “From Genus to Phylum: LSU and ITS rRNA operon regions showed similar classification accuracy influenced by database composition”, *Applied and Environmental Microbiology*, 2014, 80(3), 829-40

T. T. Wong, **K. L. Liu** and Y.D. Han “Dirichlet priors for Markov naïve Bayesian classifiers with Multinomial model for gene sequence data”, *Journal of Information Management*, 2015, Vol. 22, No. 1

- Conference Paper

K. L. Liu, T. T. Wong, G. Xie, N. W. Hengartner, 2009 July, Improving Naïve Bayesian Classifier for Metagenomic reads assignment, *In Proceedings of the International Conference on Bioinformatics & Computational Biology (BIOCOMP 2009)*, Las Vegas, U.S.A.

K. L. Liu, T. T. Wong, 2009 April, Genome signature based on Bayesian Markov model, *The 20th International Conference on Information Management (ICIM 2009)*, Taipei, Taiwan.

C. V. Trappey, H.Y. Wu, **K. L. Liu**, Knowledge discovery of customer satisfaction and dissatisfaction using ontology-based text analysis of critical incident dialogues. *CSCWD 2012*: 470-475

C. V. Trappey, H.Y. Wu, **K. L. Liu**, F. T. Lin, Knowledge Discovery of Service Satisfaction Based on Text Analysis of Critical Incident Dialogues and Clustering Methods. *ICEBE 2013*: 265-270

H. Y. Wu, **K. L. Liu** and C. V. Trappey, Understanding customers using Facebook Pages: Data mining users feedback using text analysis. *CSCWD 2014*: 346-350

- Oral and Poster presentation

K. L. Liu, C. Kuske, G. Xie, N. Hengartner, P.S.G. Chain, 2011 “Accurate taxonomy assignments from 16S illumine tags”, *Human Microbiome Research Conference*, St. Louis, U.S.A.

K. L. Liu, G. Xie, A. Porras-Alfaro, S. Eichorst, N. W. Hengartner, C. R. Kuske, 2011 “Curated LSU rRNA and ITS datasets and a fungal classifier for short read sequences”, *13th International Symposium on Microbial Ecology*, Seattle, U.S.A.

K. L. Liu, G. Xie, C. R. Kuske, P. Chain and N. W. Hengartner, “16S rRNA sequence taxonomy assignment by repetitive oligonucleotide”, 2010 IDeA Network for Biomedical Research Excellence (INBRE), Santa Fe, U.S.A.

K. L., Liu, G. Xie and N. W. Hengartner, “Improving Naïve Bayesian Classifier for Metagenomic reads assignment”, 2009 IDeA Network for Biomedical Research Excellence (INBRE), Santa Fe, U.S.A.

J. Dunbar, J. Gans, A Porras-Alfaro, G. Xie, **K. L. Liu**, C. Hesse, L. V. Gallegos-Graves, and C. R. Kuske, 2014 “Molecular tools and databases to monitor and classify soil fungal and bacterial communities”, Department of Energy’s annual Genomic Science Contractors-Grantees Meeting, Feb 9-12, Arlington, U.S.A.

A Porras-Alfaro, **K. L. Liu**, Z. Gossage, L. Johnson, T. Williams, G. Xie, and C. R. Kuske. 2012. ITS and LSU automated classification: the fungal RDP naïve Bayesian classifier. Mycological Society of America, July 14-19, Yale University, New Haven.

K. L. Liu, G. Xie, S. Eichorst, N. Hengartner, C.R. Kuske, A. Porras-Alfaro, 2011. Automated classification of fungal communities. LTER Fungal/Microbial Genomics Working Group Meeting. Long-Term Ecological Research Network. June 16-18th.

TEACHING

Teaching Assistant

Feb, 2014 ~ Jan, 2015

Inst. of Information Management, National Cheng Kung University

- Advisor: Dr. Tzu-Tsung, Wong
- Subject: Statistics I & II

Teaching Assistant

Feb. 2011 ~ Jun. 2011

Inst. of Information Management, National Cheng Kung University

- Advisor: Dr. Tzu-Tsung, Wong
- Subject: Database Management
- <http://140.116.5.200/~r7896103/DB>

Teaching Assistant

Feb. 2008 ~ Jun. 2008

Inst. of Information Management, National Cheng Kung University, Taiwan

- Advisor: Dr. Yeu-Shiang Huang
- Subject: Statistics II
- <http://140.116.5.200/~r7896103/statistics>

Teaching Assistant for foreign national students

July. 2007 ~ Apr. 2008

National Cheng Kung University, Taiwan

- Subject: Statistics, Management Science

Department of Information Management, Shih Chien University

- Subject: Computer Organization
- <http://140.116.5.200/~r7896103/ca>

Teaching Assistant

Sep. 2007~Jan.2008

Inst. of Information Management, National Cheng Kung University, Taiwan

- Advisor: Dr. Tzu-Tsung, Wong
- Subject: Statistics I

Teaching Assistant

Feb. 2006 ~ Jun. 2006

Inst. of Information Management, National Cheng Kung University, Taiwan

- Advisor: Dr. Tzu-Tsung, Wong
- Subject: Database Management System

REFERENCES

Dr. Tzu-Tsung Wong (Dissertation advisor)

Professor

Institute of Information Management National Cheng Kung University Tainan, Taiwan

Tel: +886(6) 2757575#53722

Email: tzutsung@mail.ncku.edu.tw

Dr. Gary Xie (Mentor of all projects at LANL)

Technical Staff Member Bioscience Division

Los Alamos National Laboratory Los Alamos, NM 87545 USA

Tel: +1(505) 667-2310 Fax: +1(505) 665-3024

Email: xie@lanl.gov

Dr. Patrick Chain (Mentor of Metagenomic projects at LANL)

Metagenomics and Bioinformatics Analytics Team Leader

Bioscience Division, Los Alamos National Laboratory Los Alamos, NM 87545 USA

Tel: +1(505) 665-4019 Fax: +1(505) 665-3024

Email: pchain@lanl.gov

Dr. Cheryl Kuske (Fungal LSU/ITS classifier project supervisor)

Laboratory Fellow

Bioscience Division Los Alamos National Laboratory Los Alamos, NM 87545 USA

Tel: +1(505) 665-4800 Fax: +1(505) 665-3024

Email: kuske@lanl.gov