

**Hsiang-Fu Yu 余相甫**

Amazon Inc.  
 Personalization Science  
 1900 University Ave. East Palo Alto, CA, 94303

rofuyu@cs.utexas.edu  
 Phone: (+1) 512-666-7638  
<http://www.cs.utexas.edu/~rofuyu>

**EDUCATION**

- Ph.D. Dept. of Computer Science, The University of Texas at Austin** Aug. 2011 - Aug. 2016  
*GPA 4.00/4.00*  
*Member of the Center of Big Data Analytics; Advisor: Professor Inderjit S. Dhillon*
- M.S. Dept. of Computer Science, National Taiwan University** Sep. 2008 - Jun. 2010  
*GPA 4.00/4.00; Ranked 1<sup>st</sup> out of 140*  
*Member of the Machine Learning and Data Mining Group; Advisor: Professor Chih-Jen Lin*
- B.S. Dept. of Computer Science, National Taiwan University** Sep. 2004 - Jun. 2008  
*GPA 4.00/4.00 (CS Major), 3.99/4.00 (Total)*  
*Ranked 2<sup>nd</sup> out of 118, with 8 Presidential Awards (top 5% each semester)*  
*Member of the Machine Learning and Data Mining Group; Advisor: Professor Chih-Jen Lin*

**INDUSTRY EXPERIENCE**

- Applied Scientist in Amazon, Palo Alto** Oct. 2016 - present  
 - Work in Personalization team
- Research Internship in Microsoft, Redmond** Jun. 2015 - Aug. 2015  
 - Worked on one-class collaborative filtering problems  
 - Investigated efficient algorithms and conducted empirical comparison among various approaches
- Consultant in Appier, Taipei** Dec. 2014 - Dec. 2015  
 - Provided machine learning consulting service
- Research Internship in WalmartLabs, San Bruno** May. 2013 - Aug. 2013  
 - Worked on large-scale demand-prediction problems  
 - Investigated matrix factorization techniques for time series data
- Research Internship in Ebay ResearchLabs, San Jose** May. 2012 - Aug. 2012  
 - Investigated large-scale product title classification  
 - Conducted analysis on the feedback given by buyers.
- Engineering Internship in Google Inc, Taipei/Mountain View** Jun. 2008 - Sep. 2008  
 - Worked on universal search project.  
 - Investigated machine learning approaches for boosting CTR of universal search results.

**RESEARCH INTERESTS**

*I am interested in large-scale machine learning, data-mining, and related topics:*

- Parallel algorithms for matrix completion/factorization
- Optimization techniques for multi-label Learning
- Scalable Inference methods for topic modeling

## SELECTED AWARDS AND HONORS

<b>Intel PhD Fellowship</b>	2014 – 2015
<b>Best Paper Award at IEEE ICDM 2012</b>	2012
<b>Best Research Paper Award at ACM KDD 2010</b>	2010
<b>First Prize at KDD Cup 2010</b>	2010
<b>Third Prize in the “Slow” Track of KDD Cup 2009</b>	2009
<b>Winner in SVM Track of Pascal Large Scale Learning Challenge in ICML 2008 Workshop</b>	2008

## REFEREED JOURNAL PUBLICATIONS

- [1] **H.-F. Yu**, C.-J. Hsieh, H. Yun, S.V.N. Vishwanathan, and I. S. Dhillon. Nomadic Computing for Big Data Analytics. *IEEE Computer*, vol. 49:4:52-60, 2016.
- [2] H. Yun, **H.-F. Yu**, C.-J. Hsieh, S.V.N. Vishwanathan, and I. S. Dhillon. NOMAD: Non-locking, stOchastic Multi-machine algorithm for Asynchronous and Decentralized matrix completion. *Proceedings of the VLDB Endowment*, 7:11:975-986, 2014.
- [3] **H.-F. Yu**, C.-J. Hsieh, S. Si, and I. S. Dhillon. Parallel Matrix Factorization for Recommender Systems. *Knowledge and Information Systems (KAIS)*, 2013.
- [4] **H.-F. Yu**, C.-J. Hsieh, K.-W. Chang, and C.-J. Lin, Large Scale Linear Classification When Data Cannot Fit In Memory. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 5:23:1-23, 2012.
- [5] **H.-F. Yu**, F.-L. Huang, and C.-J. Lin. Dual coordinate descent methods for logistic regression and maximum entropy models. *Machine Learning*, 85:41-75, 2011.

## REFEREED CONFERENCE PUBLICATIONS

- [6] **H.-F. Yu**, M. Bilenko, C.-J. Lin. Selection of negative samples for one-class matrix factorization. In *SIAM International Conference on Data Mining (SDM)*, 2017.
- [7] **H.-F. Yu**, H.-Y. Huang, I. S. Dhillon, and C.-J. Lin. A Unified Algorithm for One-class Structured Matrix Factorization with Side Information. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2017
- [8] **H.-F. Yu**, N. Rao, I. S. Dhillon. Temporal Regularized Matrix Factorization for High-dimensional Time Series Prediction. In *Advances in Neural Information Processing Systems (NIPS)*, 2016.
- [9] Y. You, X. Lian, J. Liu, **H.-F. Yu**, I. S. Dhillon, J. Demmel, and C.-J. Hsieh. Asynchronous Parallel Greedy Coordinate Descent. In *Advances in Neural Information Processing Systems (NIPS)*, 2016.
- [10] N. Rao, **H.-F. Yu**, P. Ravikumar, I. S. Dhillon. Collaborative Filtering with Graph Information: Consistency and Scalable Methods. In *Advances in Neural Information Processing Systems (NIPS)*, 2015.
- [11] C.-J. Hsieh, **H.-F. Yu**, I. S. Dhillon. PASSCoDe:Parallel ASynchronous Stochastic dual Co-ordinate Descent. In *International Conference Machine Learning (ICML)*, 2015.
- [12] **H.-F. Yu**, C.-J. Hsieh, H. Yun, S.V.N. Vishwanathan, and I. S. Dhillon. A Scalable Asynchronous Distributed Algorithm for Topic Modeling. In *ACM WWW International conference on World Wide Web*, 2015.
- [13] **H.-F. Yu**, P. Jain, P. Kar, and I. S. Dhillon. Large-scale Multi-label Learning with Missing Labels. In *International Conference Machine Learning (ICML)*, 2014

- [14] **H.-F. Yu**, C.-J. Hsieh, S. Si, and I. S. Dhillon. Scalable Coordinate Descent Approaches to Parallel Matrix Factorization for Recommender Systems. In *IEEE International Conference on Data Mining (ICDM)*, 2012. **Best Paper Award**.
- [15] **H.-F. Yu**, C.-J. Hsieh, K.-W. Chang, and C.-J. Lin. Large Scale Linear Classification When Data Cannot Fit In Memory. In *International Joint Conferences on Artificial Intelligence (IJCAI)*, 2011. **The Best Paper Track**.
- [16] **H.-F. Yu**, C.-J. Hsieh, K.-W. Chang, and C.-J. Lin. Large Scale Linear Classification When Data Cannot Fit In Memory. In *ACM SIGKDD International Conference of Knowledge Discovery and Data Mining (KDD)*, 2010. **Best Research Paper Award**.

## OTHER PUBLICATIONS

- [17] **H.-F. Yu**, N. Rao, and I. S. Dhillon. High dimensional time series prediction with missing values. In ArXiv, 2016
- [18] **H.-F. Yu**, N. Rao, and I. S. Dhillon. Temporal Regularized Matrix Factorization. In *Time Series Workshop of NIPS*, 2015.
- [19] **H.-F. Yu**, C.-H. Ho, Y.-C. Juan, C.-J. Lin. LibShortText: a library for short-text classification and analysis. Technical report, 2013.
- [20] **H.-F. Yu**, C.-H. Ho, P. Arunachalam, M. Somaiya, and C.-J. Lin. Product title classification versus text classification. Technical report, 2012.
- [21] **H.-F. Yu**, H.-Y. Lo, H.-P. Hsieh, J.-K. Lou, T McKenzie, J.-W. Chou, P.-H. Chung, C.-H. Ho, C.-F. Chang, Y.-H. Wei, J.-Y. Weng, E.-S. Yan, C.-W. Chang, T.-T. Kuo, Y.-C. Lo, P.-T. Chang, C. Po, C.-Y. Wang, Y.-H. Huang, CW Hung, YX Ruan, Y.-S. Lin, S.-D. Lin, H.-T. Lin and C.-J. Lin. Feature engineering and classifier ensemble for KDD Cup 2010. (**First Prizes** of Both All Teams and All Student Teams).
- [22] H.-Y. Lo, K.-W. Chang, S.-T. Chen, T.-H. Chiang, C.-S. Ferng, C.-J. Hsieh, Y.-K. Ko, T.-T. Kuo, H.-C. Lai, K.-Y. Lin, C.-H. Wang, **H.-F. Yu**, C.-J. Lin, H.-T. Lin and S.-d. Lin. An Ensemble of Three Classifiers for KDD Cup 2009: Expanded Linear Model, Heterogeneous Boosting, and Selective Naive Bayes. In G. Dror et al., eds., *Proceedings of KDD-Cup 2009 competition, vol. 7 of JMLR Workshop and Conference Proceedings*, 57-64, 2009. (**Third Place** of the Slow Track).
- [23] **H.-F. Yu**, C.-J. Hsieh, K.-W. Chang, and C.-J. Lin. Pascal challenge: Linear support vector machines. In *Pascal Large Scale Learning Challenge in ICML Workshop*, 2008.

## SOFTWARE DESIGN

### **LIBPMF - A Library for Large-scale Parallel Matrix Factorization**

<http://www.cs.utexas.edu/~rofuyu/libpmf>

- Developed fast and scalable method CCD++ for matrix factorization
- The library follows the similar interface LIBSVM/LIBLINEAR, which is simple to use.
- Developed interfaces to support various languages (Python, Matlab, R).

### **LibShortText - A Library for Short-text Classification and Analysis**

<http://www.csie.ntu.edu.tw/~cjlin/libshorttext/>

- Developed an easy-to-use library which is more efficient than general text-mining packages.
- Studied carefully to select appropriate default options which works well in most situations without tedious tuning.

- Designed an iterative for error analysis.

### **LIBLINEAR - A Library for Large Linear Classification**

<http://www.csie.ntu.edu.tw/~cjlin/liblinear>

- Developed a new solver for LIBLINEAR.
- Developing/maintaining the library and answering questions from users.
- The library has been used in some major Internet companies to classify their web data

### **LIBSVM - A Library for Support Vector Machine**

<http://www.csie.ntu.edu.tw/~cjlin/libsvm>

- Developed a new python interface for LIBSVM
- Developing/maintaining the library and answering questions from users
- The library has been downloaded more than 150,000 times since Apr. 2000

## **OTHER AWARDS AND HONORS**

<b>MCD Fellowship</b>	2011 – 2013
<b>Master's Thesis Award, Taiwanese Association for Artificial Intelligence</b>	2010
<b>The Scholarship of Pan Wen Yuan Foundation</b>	2009
<b>Honorary Member of the Phi Tau Phi Scholastic Honor Society</b>	2008
officially recommended by NTU, from top 2% of 118 graduating students in computer science department	
<b>The Scholarship of CyberLink</b>	2007
<b>ACM ICPC Asia Regional programming Contest</b>	
Eighth Place, Amritapuri	2007
Thirteen Place, Taipei	2007
Bronze Medal, Taipei/Xian	2006
<b>Presidential Award, National Taiwan University</b>	Fall/Spring 2004 - 2008
given to the top 5% undergraduate students each semester	

## **SKILLS**

- Programming: C/C++, Python, MATLAB, OpenMP, MPI