

Keke Chen

Phone: (office) +1 (937) 775-4642, (cell) +1 (937) 212-5919
E-Mail: keke.chen@wright.edu
WWW: <http://www.cs.wright.edu/~keke.chen>

RESEARCH INTERESTS

- Big Data in the Cloud: Elastic and Economical Data Intensive Processing, Scalable Algorithms, Visual Analytics, and Privacy-Preserving Outsourced Computation;
- Privacy Issues in Social Computing;
- Core Machine Learning and Data Mining Algorithms: Learning to Rank, Domain Adaptation Algorithms, Clustering Algorithms, Clustering Analysis, and Applications in Bioinformatics.

EDUCATION

Georgia Institute of Technology, Atlanta GA	Ph.D. in Computer Science	(9/2001 – 8/2006)
Advisor: Professor Ling Liu; Research areas: data privacy, data mining, distributed and parallel computing Minor: Statistics		
Zhejiang University, China	M.Sc. in Computer Science	(9/1996 – 4/1999)
Tongji University, China	B.Sc. Degree in Computer Science	(9/1992 – 7/1996)

PROFESSIONAL EXPERIENCE

Visiting Scholar, Bioinformatics for Immunology, University of Texas at Austin	(9/2015 – 5/2016)
Director of Data Intensive Analysis and Computing (DIAC) Lab	(9/2008-present)
Associate Professor, Computer Science and Engineering, Wright State University,	(8/2014 – present)
Assistant Professor, Computer Science and Engineering, Wright State University,	(8/2008 – 8/2014)
Senior Research Scientist, Yahoo! Labs, Santa Clara CA	(8/2006 – 8/2008)
<i>Learning to rank from pairwise training data (GBRank); ranking function adaptation (Trada); user preference mining; big data analytics with Hadoop</i>	
Graduate Research Assistant, Georgia Institute of Technology, Atlanta GA	(5/2002 – 8/2006)
Research Intern, Yahoo! Search, Sunnyvale CA	(Summers 2004 and 2005)
Research Scientist/Lecturer, Zhejiang University, China	(5/1999 – 8/2000)

RESEARCH PROJECTS

- **Big Data in the Cloud**
 - **CloudVista**: Interactive and economical visual data analytics for big data in the cloud. [SSDBM11 paper](#), [VLDB12 demo](#), [CLOUD 15 paper](#) (collaborator: Dr. Pak C. Wong at Pacific Northwest National Laboratory (PNNL))
 - **CRESP**: Cloud resource provisioning for large scale MapReduce programs. [TPDS14 paper](#), [CLOUD11 paper](#).
 - **RASP**: random space perturbation for privacy preserving outsourced query services and data mining. [CODASPY 11 paper](#), [CCS12 poster](#), [ICDM13paper](#), [TKDE12 paper](#), [VLDB14demo](#)
 - **Geometric data perturbation** for privacy-preserving outsourced data mining, [KAIS10 paper](#)
 - Secure computation and mining of big data in the cloud: the dense matrix and sparse matrix approaches. [CLOUD13 paper](#), [ASIACCS16 paper](#)
 - Big data applications:
 - Large-scale sequencing data and gene interaction network analysis for immunology research (collaborator: Prof. Ning Jiang at University of Texas at Austin),
 - Image stream analysis and querying for space situational awareness (collaborator: Dr. Vincent Schmidt at Air Force Research Lab (AFRL)).

- **Privacy Issues with Social Computing:**
 - [PUTS](#): Privacy-utility tradeoffs in social network privacy settings. [PASSAT12 paper](#)
 - Mining and understanding regrettable tweets for privacy protection. [WWW15 poster](#), [WWW16 paper](#)

- **Learning to Rank and Adapting to Rank:**
 - **Trada**: regression tree based domain adaptation, [ACM TOIS 11 paper](#), [CIKM08 paper](#), [patent](#)
 - Domain Similarity Analysis for Web Search Ranking, [CIKM09 paper](#)
 - Visual Analysis of Regression Trees for Ranking, [CollaborateCom09 paper](#)
 - Tradeoffs on User Preferences and Expert Judgment for Learning Ranking Functions, [DBRANK08 Paper](#), [Patent](#)
 - **GBRank**: Learning to rank with pairwise preference data in the regression framework, [SIGIR07 paper](#), [NIPS07 paper](#), [patent](#)

- **Clustering Analysis and Visualization:**
 - [BestK: Categorical data clustering and validation](#)
 - **TransCluster**: [Transactional data clustering](#) and [validation](#)
 - [Stream BestK](#): Change detection on categorical/transactional data streams
 - [iVIBRATE: a visual framework for clustering large datasets](#)
 - [VISTA: visual rendering and validating clustering structures](#)

TEACHING

- CS790 Privacy Aware Computing (Winter2009, Winter2011, Winter 2012)
- CS7850 Privacy Aware Computing (Spring 2013, Spring 2014)
- CEG435/635 Distributed Computing and Systems (Winter2010)
- CEG4360/6360 Distributed Systems and Cloud Computing (Spring2013, Spring2014)
- CS499/699 Cloud Computing (Fall2009, Fall2010, Fall2011, Spring2012)
- CEG7380 Cloud Computing (Fall2012, Fall 2013, Fall 2014)
- CS405/605 Introduction to Database Management Systems (Spring2009, Fall2009, Spring2010, Fall2010, Winter2011, Spring2011, Spring2012)
- CS240 Introduction to Computer Science I: Java Programming (Spring 2011)
- CS400/600 Data Structures and Algorithms (Fall2008, Fall2011, Winter2012)
- CS3100/5100 Data Structures and Algorithms (Fall 2012, Spring 2013)

- Developed new courses: “[Cloud Computing](#)”, “[Privacy Aware Computing](#)”, and “Distributed Systems and Cloud Computing”

PUBLICATIONS (www.cs.wright.edu/~keke.chen/pubs.html for the full list)

Journal Papers:

1. James Powers, Sagar Sharma and **Keke Chen**, “PEig: Privacy-Preserving Eigendecomposition of Outsourced Matrices in the Cloud”, submitted to IEEE Transactions on Dependable and Secure Computing.

2. **Keke Chen** and Shumin Guo, “RASP-Boost: Confidential Boosting-Model Learning with Perturbed Data in the Cloud”, IEEE Transactions on Cloud Computing, 2015.

3. **Keke Chen**, James Powers, Shumin Guo, and Fengguang Tian, "CRESP: Towards Optimal Resource Provisioning for MapReduce Computing in Public Clouds", IEEE Transactions on Distributed and Parallel Systems (TPDS), Volume 25, Number 6, 2014
4. Huiqi Xu, Shumin Guo, and **Keke Chen**, "Building Confidential and Efficient Query Services in the Cloud with RASP Data Perturbation ", IEEE Transactions on Knowledge and Data Engineering (TKDE), Volume 26, Number 2, 2014
5. **Keke Chen**, "Optimizing Star-coordinate Visualization Models for Effective Interactive Cluster Exploration on Big Data ", Journal of Intelligent Data Analysis, Volume 18, Number 2, 2014
6. **Keke Chen**, Jing Bai and Zhaohui Zheng, "Ranking Function Adaptation with Boosting Trees", ACM Transactions on Information Systems, Vol. 29, No. 4, 2011
7. **Keke Chen** and Ling Liu, "Geometric Data Perturbation for Privacy Preserving Outsourced Data Mining", Journal of Knowledge and Information Systems (KAIS), Vol. 29, No. 3, 2011
8. Hua Yan, **Keke Chen** and Ling Liu, "*SCALE: a Scalable Framework for Efficiently Clustering Large Transactional Data* ", Journal of Data Mining and Knowledge Discovery (DMKD), Vol. 20, No. 1, 2010. (corresponding author)
9. **Keke Chen** and Ling Liu, "*Privacy-preserving Multiparty Collaborative Mining with Geometric Data Perturbation*", IEEE Transactions on Parallel and Distributed Systems, Vol. 20, No.12, 2009
10. **Keke Chen** and Ling Liu, "*HE-Tree: a Framework for Detecting Changes in Clustering Structure for Categorical Data Streams* ", VLDB Journal, Vol. 18, No. 6, 2009
11. Hua Yan, **Keke Chen** and Ling Liu, "*Determining the Best K for Clustering Transactional Datasets: A Coverage Density-based Approach*", Journal of Data and Knowledge Engineering (DKE), Vol. 68, No.1, 2009
12. **Keke Chen** and Ling Liu, "*Best K: the Critical Clustering Structure in Categorical Dataset*", Journal of Knowledge and Information Systems (KAIS), Vol. 20, No. 1, 2009
13. Yongjing Lin, Wenyan Li, **Keke Chen** and Ying Liu, "*Document Clustering and Ranking System for Exploring MEDLINE Citations*", Journal of the American Medical Informatics Association (JAMIA), Volume 14, Issue 5, 2007.
14. **Keke Chen** and Ling Liu, "iVIBRATE: Interactive Visualization Based Framework for Clustering Large Datasets " ACM Transactions on Information Systems (TOIS), Vol24, issue 2, 2006
15. **Keke Chen** and Ling Liu, "VISTA: Validating and Refining Clusters via Visualization." Journal of Information Visualization, Volume 3, number 4, Sept. 2004

Conference Papers:

16. Sagar Sharma, James Powers, and **Keke Chen**, "Privacy-Preserving Spectral Analysis of Large Graphs in Public Clouds", ACM Symposium on Information, Computer and Communications Security (ASIACCS), Xi'an, China, 2016
17. Lu Zhou, Wenbo Wang, and **Keke Chen**, "Tweet Properly: Analyzing Deleted Tweets to Understand and Identify Regrettable ones", International conference on World Wide Web (WWW), Montreal, Canada, 2016
18. Lu Zhou, Wenbo Wang, and **Keke Chen**, "Identifying Regrettable Messages from Tweets", International conference on World Wide Web (WWW), poster session, Florence, Italy, 2015

19. Zohreh Alavi and **Keke Chen**, "Scaling Up Euclidean Embedding Algorithms for Big Data", IEEE Cloud Computing conference (CLOUD), New York City, 2015
20. Zohreh Alavi, Lu Zhou, James Powers, and **Keke Chen**, "RASP-QS: Efficient and Confidential Query Services in the Cloud", Demonstration Session of Very Large Databases (VLDB), Hangzhou, China, 2014
21. **Keke Chen** and Shumin Guo, "PerturBoost: Practical Confidential Classifier Learning in the Cloud", IEEE Conference on Data Mining (ICDM), Dallas, TX 2013
22. James Powers and **Keke Chen**, "Secure Computation of Top K Eigenvectors for Shared Matrices in the Cloud", IEEE Cloud Computing conference (CLOUD), Santa Clara, CA 2013.
23. Shumin Guo and **Keke Chen**, "Privacy Preserving Boosting in the Cloud with Secure Half-Space Queries", Poster Session of ACM Conference on Computer and Communication Security (CCS), 2012
24. Shumin Guo and **Keke Chen**, "Mining Privacy Settings to Find Optimal Privacy-Utility Tradeoffs for Social Network Services", ASE/IEEE International Conference on Privacy, Security, Risk and Trust (PASSAT), 2012
25. Huiqi Xu, Zhen Li, Shumin Guo, and **Keke Chen**, "CloudVista: Interactive and Economical Visual Cluster Analysis for Big Data in the Cloud". Demonstration Session of International Conferences on Very Large Databases, Istanbul, Turkey 2012.
26. Jinpeng Wei, Calton Pu, and **Keke Chen**, "Flying under the Radar: Maintaining Control of Kernel without Changing Kernel Code or Persistent Data Structures". The 7th Annual Cyber Security and Information Intelligence Research Workshop (CSIIRW), October 12 - 14, 2011.
27. **Keke Chen**, Ramakanth Kavuluru, and Shumin Guo, "RASP: Efficient Multidimensional Range Query on Attack-Resilient Encrypted Databases", ACM Conference on Data and Application Security and Privacy (CODASPY), San Antonio, TX, 2011
28. **Keke Chen**, Huiqi Xu, Fengguang Tian, and Shumin Guo, " CloudVista: visual cluster exploration for extreme scale data in the cloud ", Scientific and Statistical Database Management Conference, Portland OR, 2011
29. Fengguang Tian and **Keke Chen**, " Towards Optimal Resource Provisioning for Running MapReduce Programs in Public Clouds", IEEE Conference on Cloud Computing, Washington DC, 2011
30. Jing Bai, Fernando Diaz, Yi Chang, Zhaohui Zheng and **Keke Chen**, " Cross-Market Model Adaptation with Pairwise Preference Data for Web Search Ranking ", International Conference on Computational Linguistics (COLING), 2010
31. **Keke Chen**, Jing Bai, Srihari Reddy, and Belle Tseng, "*On Domain Similarity and Effectiveness of Adapting to Rank* ", ACM Conference on Information and Knowledge Management (CIKM09), 2009.
32. **Keke Chen** and Fengguang Tian, " VisGBT: Visually Analyzing Evolving Datasets for Adaptive Learning ", The 5th International Conference on Collaborative Computing (CollaborateCom09), 2009.
33. **Keke Chen**, Rongqing Lu, CK Wong, Gordon Sun, Larry Heck and Belle Tseng, "*Trada: Tree-Based Ranking Function Adaptation*", ACM Conference on Information and Knowledge Management (CIKM), Oct. 2008
34. **Keke Chen**, Ya Zhang, Zhaohui Zheng, Hongyuan Zha, and Gordon Sun, "*Adapting Ranking Functions to User Preferences*", DBRank workshop at IEEE conference on Data Engineering (ICDE), 2008

35. **Keke Chen** and Ling Liu, *Towards Attack-Resilient Geometric Data Perturbation*, in Proceedings of the SIAM Data Mining Conference (SDM07), Minneapolis, MN, April 2007.
36. Zhaohui Zheng, **Keke Chen**, Gordon Sun, and Hongyuan Zha, *A Regression Framework for Learning Ranking Functions Using Relative Preferences*, in Proceedings of ACM SIGIR Conference, Amsterdam, Netherlands, July 2007
37. Zhaohui Zheng, Hongyuan Zha, Tong Zhang, Olivier Chapelle, **Keke Chen**, and Gordon Sun, *A General Boosting Method and its Application to Learning Ranking Functions for Web Search*, in Proceedings of Neural Information Processing Systems (NIPS), Canada, 2007
38. **Keke Chen** and Ling liu, "Space Adaptation: Privacy-preserving Multiparty Collaborative Mining with Geometric Perturbation ", ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing (PODC 2007) (Brief Announcement)
39. **Keke Chen** and Ling Liu, "Detecting Change of Clustering Structure in Categorical Data Stream ", SIAM International Conference on Data Mining Conference(SDM), 2006
40. Hua Yan, **Keke Chen** and Ling Liu, " Efficiently Clustering Transactional Data with Weighted Coverage Density", ACM Conference on Information and Knowledge Management (CIKM2006) , November, 2006
41. **Keke Chen** and Ling Liu, "Privacy-Preserving Data Classification with Rotation Perturbation ", Proc. of IEEE Intl. Conf on Data Mining 2005 (ICDM05)
42. **Keke Chen** and Ling Liu, "The 'Best K' for Entropy-based Categorical Clustering ", Proc of Scientific and Statistical Database Management (SSDBM05). Santa Barbara, CA June 2005.
43. **Keke Chen** and Ling Liu, "ClusterMap: Labeling Large Datasets via Visualization." ACM Conf. of Information and Knowledge Management (CIKM04), Washington DC, Nov, 2004
44. **Keke Chen** and Ling Liu, "Validating and Refining Clusters via Visual Rendering." Proc. of Intl. Conf. on Data Mining(ICDM03). Melbourne, FL, November 2003.
45. **Keke Chen** and Ling Liu, "A Visual Framework Invites Human into the Clustering Process." Proc. of Scientific and Statistical Database Management (SSDBM03). Cambridge, Boston, July 2003.
46. **Keke Chen** and Ling Liu, "Cluster Rendering of Skewed Datasets via Visualization." Proc. of ACM Symposium on Applied Computing(ACM SAC03). Melbourne, FL, March 2003

Book Chapters

47. **Keke Chen** and Ling Liu, "A Survey of Multiplicative Data Perturbation for Privacy Preserving Data Mining." In "Privacy-Preserving Data Mining: Models and Algorithms", edited by Charu C. Aggarwal and Philip S. Yu, 2008
48. **Keke Chen**, Shumin Guo, James Power, Fengguang Tian, "Optimal Resource Provisioning for Economical and Green MapReduce Computing in the Cloud", to appear in the book "Data Processing Techniques in the Era of Big Data", edited by Sherif Sakr and Mohamed Medhat, CRC Press, 2013.

PATENTS

-
- [PT.1] **Keke Chen**, Ya Zhang, Zhaohui Zheng, Hongyuan Zha, and Gordon Sun: Systems and Apparatus for Using B Measures to Learn Balanced Relevance Functions from Expert and User Judgments, US8126839, filing date: May 23, 2007, issue date: Mar 30, 2010

[PT.2] **Keke Chen**, Rongqing Lu, CK Wong, Gordon Sun, Larry Heck, and Belle Tseng: Methods and Apparatuses for Adapting a Ranking Function of a Search Engine for Use with a Specific Domain, US7685078, filing date: Jun 19, 2008, issue date: Feb 28 2012

[PT.3] Zhaohui Zheng, Hongyuan Zha, **Keke Chen**, and Gordon Sun: A Regression Framework for Learning Ranking Functions Using Relative Preferences, US20080208836, filing date: Feb 23, 2007, issue date: August 28, 2008

PROFESSIONAL ACTIVITIES

Honors for Professional Services:

- Excellent Service at International Conference on Collaborative Computing, 2009
- Diligent PC Member of IEEE CLOUD Conferences, 2014

Editorial Board: International Journal of Cloud Computing (<http://www.hipore.com/ijcc/>)

Publicity Chair: On the Move Federated Conferences and Workshops (OTM2008)

Tutorial Chair: International Conference on Collaborative Computing (CollaborateCom 2009)

Area Chair: Data Analytics in the Cloud, IEEE CLOUD 2016

Program Committee Member:

IEEE International Conference on Big Data (IEEE BigData 2013, 2014, 2015)

IEEE International Conference on Big Data Science and Engineering (BigDataSE 2015, 2016)

IEEE International Conference on Big Data and Distributed Systems (2012, 2013, 2014)

International Symposium on Big Data Management and Analytics (2016)

IEEE International Conference on Cloud and Service Computing (CSC2011, 2012, 2013, 2014, 2015)

IEEE International Conference on Bioinformatics and Bioengineering, (BIBE 2014)

IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2011, 2012, 2014, 2015)

The International Conference on Web Information System Engineering (WISE 2011)

The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD 2011, 2014, 2015)

IEEE Conference on Cloud Computing (CLOUD 2011, 2012, 2013, 2014, 2015)

IEEE Conference on Distributed Computing Systems (ICDCS 2011, 2012)

IEEE Conference on Web Services (ICWS 2010, 2011, 2012)

IEEE Conference on Data Mining (ICDM2008)

International Future Internet Symposium (FIS2008, 2009, 2012),

International Conference on Cooperative Information Systems (CoopIS 2008)

International Conference on Cloud Computing and Big Data (CCBD 2014)

International Workshop on Management of Cloud Systems (MoCS 2012, 2013, 2014)

International Workshop on Collaborative Cloud (CollabCloud 2014)

MobiCloud Workshop in Conjunction with MobiCASE 2010

International Workshop on Collaborative Big Data (C-Big 2012, 2013, 2014)

International Workshop on Privacy, Security and Trust in Mobile and Wireless Systems (MobiPST 2011, 2012, 2013, 2014, 2015)

International Workshop on Machine learning, Optimization and big Data (MOD 2015)

International Symposium on Big Data and Data Analytics in Collaboration (BDDAC 2015)

International Workshop on SECURITY, PRIVACY, and TRUSTWORTHINESS IN MEDICAL CYBER PHYSICAL SYSTEM (MedSPT 2016)

Reviewer for Professional Journals

SIGMOD Record

Knowledge and Information Systems (KAIS)

Journal of Data Mining and Knowledge Discovery (DMKD)

IEEE Transactions on Knowledge and Data Engineering (TKDE)

IEEE Internet Computing

Pattern Recognition
IEEE Transactions on Neural Network
Journal of Enterprise Information Systems
VLDB Journal
ACM Transactions on Knowledge Discovery from Data (TKDD)
Journal of Parallel and Distributed Computing
IEEE Transactions on Dependable and Secure Computing
ACM Transactions on Information Systems
Journal of Computer Security
Transactions on Data Privacy