# Shipeng Yu

Staff Scientist

Siemens Medical Solutions USA, Inc.

## CONTACT INFORMATION

Tel:      +1-610-448-4420 (O)

Fax:      +1-610-448-4274 (O)

URL:      http://www.dbs.ifi.lmu.de/~spyu/

Email:    shipeng.yu@siemens.com (work), shipeng.yu@gmail.com (private)

## RESEARCH INTERESTS

- ✧ Statistical Machine Learning, Probabilistic Modeling, Bayesian Data Analysis
- ✧ Bayesian Clustering, Dimensionality Reduction, Supervised Projection
- ✧ Document Modeling, Collaborative Filtering, Ranking
- ✧ Information Retrieval and Extraction, Text Mining, Web Search and Mining

## EDUCATION

Ph.D.  in Institute for Computer Science, University of Munich (9/2003 - 9/2006)
      Supervisor: Prof. Dr. Hans-Peter Kriegel, Dr. Volker Tresp

M.Sc.  in Information Science, School of Mathematical Sciences, Peking University, China (9/2000 - 7/2003)
      Supervisor: Prof. Dr. Zuoquan Lin
      GPA: **91.6/100**                                    Rank: **2/70**

B.Sc.  in Information Science, School of Mathematical Sciences, Peking University, China (9/1996 - 7/2000)
      GPA: **93.5/100** for major, **91.7/100** for all              Rank: **2/41**

## RESEARCH EXPERIENCE

11/2006 – Present: **Staff Scientist** at Siemens Medical Solutions USA, Inc.

  *Directions: statistical machine learning for medical data mining*

9/2003 – 9/2006: **Guest Research Scientist** at Siemens Corporate Technology, Munich, Germany

  *Directions: statistical machine learning, data mining and information retrieval*

- ✧ Proposed a Dirichlet enhanced latent semantic analysis for documents
- ✧ Introduced a variational Bayesian method for infinite mixture of exponential family distributions
- ✧ Co-invented a supervised projection model and its probabilistic extensions and interpretations
- ✧ Co-invented a soft clustering algorithm on graph with applications on semi-supervised learning
- ✧ Introduced a multi-task Gaussian Process model for collaborative ranking problem
- ✧ Proposed a supervised and a semi-supervised PCA which can be applied to very large data sets

7/2005 - 9/2005: **Research Internship** at Microsoft Research Cambridge, Cambridge, UK

  *Topic: Bayesian web ranking with TrueSkill$^{TM}$ System*

- ✧ Addressed web ranking problem by extending TrueSkill$^{TM}$ system for games
- ✧ Considered time-independent ranking and feature-dependent ranking for web ranking scenario

- ✧ Implemented the whole algorithm to handle millions of document-query pairs
- ✧ Proposed a probabilistic method to evaluate judges and measure query difficulties

**9/2001 - 8/2003: Visiting Student** at <u>Microsoft Research Asia, Beijing, China</u>

*Directions: web page segmentation and applications on web information retrieval, extraction and mining*

- ✧ Proposed a novel vision-based content structuring method for web page
- ✧ Verified that vision-based page segmentation (VIPS) could significantly improve web IR performance
- ✧ Proposed a combined page segmentation approach which achieved the best performance
- ✧ Introduced a new strategy for automatic web information extraction based on web content structure

**1/2000 - 7/2003: Research Assistant** at <u>PKU-Goldtech Networking R&D Center, Beijing, China</u>

*Directions: web-based DRP systems with security issues*

- ✧ Team Member of *WebDaemon*, a role-based access control (RBAC) system on the web
- ✧ Project Leader of *TYGXC*, an Internet-based DRP system for TianYin Corporation, China
- ✧ Team Member of *TCLJXC*, an Internet-based DRP system for TCL Corporation, China

## SELECTED PUBLICATIONS

- ✧ Shipeng Yu, Balaji Krishnapuram, Romer Rosales, Harald Steck, and R. Bharat Rao. **Bayesian Multi-View Learning**. To appear as spotlight in the *Advances in Neural Information Processing Systems 20* (NIPS'2007), 2007.
- ✧ Jianwu Xu, Shipeng Yu, Jinbo Bi, Lucian Lita, Stefan Niculescu, and R. Bharat Rao. **Automatic Medical Coding of Patient Records via Weighted Ridge Regression**. To appear in the *6th International Conference on Machine Learning and Applications* (ICMLA'2007), 2007.
- ✧ Zhao Xu, Volker Tresp, Shipeng Yu, Kai Yu and Hans-Peter Kriegel. **Fast Inference in Infinite Hidden Relational Models**. In *the 5th International Workshop on Mining and Learning with Graphs* (MLG'2007), Florence, Italy, 2007.
- ✧ Shipeng Yu, Volker Tresp, and Kai Yu. **Robust Multi-Task learning with t-Processes**. In *the 24th International Conference on Machine Learning* (ICML'2007), Corvallis, OR, 2007.
- ✧ Mingrui Wu, Kai Yu, Shipeng Yu, and Bernhard Schölkopf. **Local Learning Projections**. In *the 24th International Conference on Machine Learning* (ICML'2007), Corvallis, OR, 2007.
- ✧ Kai Yu, Wei Chu, Shipeng Yu, Volker Tresp, and Zhao Xu. **Stochastic Relational Models for Discriminative Link Prediction**. In *Advances in Neural Information Processing Systems 19* (NIPS'2006), 2006.
- ✧ Shipeng Yu, Kai Yu, Volker Tresp, and Hans-Peter Kriegel. **Multi-Output Regularized Feature Projection**. In *IEEE Transactions on Knowledge and Data Engineering*, Vol. 18, no. 22, pp. 1600-1613, December 2006.
- ✧ Shipeng Yu, Kai Yu, Volker Tresp, and Hans-Peter Kriegel. **Variational Bayesian Dirichlet-Multinomial Allocation for Exponential Family Mixtures**. In *the 17th European Conference on Machine Learning* (ECML'2006), Berlin, Germany, 2006.
- ✧ Shipeng Yu, Kai Yu, Volker Tresp, Hans-Peter Kriegel, and Mingrui Wu. **Supervised Probabilistic Principal Component Analysis**. In *the 12th ACM International Conference on Knowledge Discovery and Data Mining* (KDD'2006), Philadelphia, PA, 2006.
- ✧ Shipeng Yu, Kai Yu, Volker Tresp, and Hans-Peter Kriegel. **Collaborative Ordinal Regression**. In *the 23rd International Conference on Machine Learning* (ICML'2006), Pittsburgh, PA, 2006.

- Shipeng Yu, Kai Yu, and Volker Tresp. **Collaborative Ordinal Regression**. In *the Proceedings of Workshop on Learning to Rank at Neural Information Processing Systems 18* (NIPS'2005), Whistler, Canada, 2005.

- Kai Yu, Shipeng Yu, and Volker Tresp. **Soft Clustering on Graphs**. *Advances in Neural Information Processing Systems 18* (NIPS'2005), Vancouver, Canada, 2005.

- Yi Huang, Kai Yu, Matthias Schubert, Shipeng Yu, and Hans-Peter Kriegel. **Hierarchy-Regularized Latent Semantic Indexing**. In *the 5th IEEE International Conference on Data Mining* (ICDM'2005), New Orleans, Louisiana, 2005.

- Shipeng Yu, Kai Yu, Volker Tresp, and Hans-Peter Kriegel. **A Probabilistic Clustering-Projection Model for Discrete Data**. In *the 9th European Conference on Principles and Practice of Knowledge Discovery in Databases* (PKDD'2005), Porto, Portugal, 2005. (**Best Paper Runner-Up / Best Student Paper Runner-Up**)

- Kai Yu, Shipeng Yu, and Volker Tresp. **Blockwise Supervised Inference on Large Graphs**. In *the Proceedings of Workshop on Learning with Partially Classified Training Data at the 22nd International Conference on Machine Learning* (ICML'2005), Bonn, Germany, 2005.

- Zhao Xu, Volker Tresp, Kai Yu, Shipeng Yu, and Hans-Peter Kriegel. **Dirichlet Enhanced Relational Learning**. In *the 22nd International Conference on Machine Learning* (ICML'2005), Bonn, Germany, 2005.

- Kai Yu, Shipeng Yu, and Volker Tresp. **Multi-Label Informed Latent Semantic Indexing**. In *the 28th Annual International ACM SIGIR Conference* (SIGIR'2005), Salvador, Brazil, 2005.

- Kai Yu, Shipeng Yu, and Volker Tresp. **Multi-Output Regularized Projection**. In *the International IEEE Conference on Computer Vision and Pattern Recognition* (CVPR'2005), San Diego, CA, 2005.

- Kai Yu, Shipeng Yu, and Volker Tresp. **Dirichlet Enhanced Latent Semantic Analysis**. In *the Tenth International Workshop on Artificial Intelligence and Statistics* (AISTATS'2005), Barbados, 2005.

- Kai Yu, Volker Tresp, and Shipeng Yu. **A Nonparametric Hierarchical Bayesian Framework for Information Filtering**. In *the 27th Annual International ACM SIGIR Conference* (SIGIR'2004), Sheffield, UK, 2004.

- Deng Cai, Shipeng Yu, Ji-Rong Wen, and Wei-Ying Ma. **Block-based Web Search**. In *the 27th Annual International ACM SIGIR Conference* (SIGIR'2004), Sheffield, UK, 2004.

- Ji-Rong Wen, Ruihua Song, Deng Cai, Kaihua Zhu, Shipeng Yu, Shaozhi Ye, and Wei-Ying Ma. **Microsoft Research Asia at the Web Track of TREC 2003**, In *the Twelfth Text Retrieval Conference* (TREC'12), 2003.

- Deng Cai, Shipeng Yu, Ji-Rong Wen, and Wei-Ying Ma. **VIPS: A Vision-based Page Segmentation Algorithm**. Microsoft Technical Report, MSR-TR-2003-79, 2003.

- Deng Cai, Shipeng Yu, Ji-Rong Wen, and Wei-Ying Ma. **Extracting Content Structure for Web Pages based on Visual Representation**. In *the Fifth Asia Pacific Web Conference* (APWeb'2003), Xi'an, China, 2003.

- Shipeng Yu, Deng Cai, Ji-Rong Wen, and Wei-Ying Ma. **Improving Pseudo-Relevance Feedback in Web Information Retrieval Using Web Page Segmentation**. In *the Twelfth International World Wide Web Conference* (WWW'2003), Budapest, Hungary, 2003.

## TALKS

- ✧ **Supervised Dimensionality Reduction with Principal Component Analysis**. Microsoft Research Redmond, Jul. 2006.
- ✧ **Learning Non-parametric Priors from Multiple Tasks**. OPEN HOUSE on Multi-Task and Complex Outputs Learning, UK, Jul. 2006.
- ✧ **Bayesian Web Ranking for Information Retrieval**. Microsoft Research Cambridge, UK, Mar. 2006.
- ✧ **Collaborative Ordinal Regression**. NIPS'05 Workshop on Learning to Rank, Dec. 2005.
- ✧ **A Probabilistic Clustering-Projection Model for Discrete Data**. The 9th European Conference on Principles and Practice of Knowledge Discovery in Databases, Porto, Portugal, Oct. 2005.
- ✧ **TrueSkill$^{TM}$ Hits the Web: Bayesian Ranking Methods for Web Search**. Intern Talk, Microsoft Research Cambridge, UK, Sep. 2005.
- ✧ **Multi-Label Informed Latent Semantic Indexing**. The 28th Annual International ACM SIGIR Conference, Salvador, Brazil, Aug. 2005.
- ✧ **Dirichlet Enhanced Latent Semantic Analysis**. Lernen, Wissensentdeckung und Adaptivität (LWA'04), Humboldt University, Berlin, Germany, Oct. 2004.
- ✧ **Enhancing Web Information Retrieval Using Web Page Segmentation**. Siemens Corporate Technology, Munich, Germany, Nov. 2003.
- ✧ **Enhancing Web Information Retrieval Using Web Page Segmentation**. Institute for Informatics, University of Munich, Munich, Germany, Oct. 2003.

## TEACHING EXPERIENCE

9/2000 - 1/2002: **Lecturer** in <u>School of Continuing Education</u>, Peking University
- ✧ *Assembly Language*, Autumn 2000 and Autumn 2001

9/2001 - 1/2003: **Teaching Assistant** in <u>School of Mathematical Sciences</u>, Peking University
- ✧ *Introduction to Database Systems*, Autumn 2002
- ✧ *MATLAB System*, Spring 2002
- ✧ *Statistical Computation*, Autumn 2001

## PATENTS

- ✧ Collaborative Ordinal Regression. Filed November 2005, Siemens Corporate Technology.
- ✧ Soft Clustering on Graphs. Filed October 2005, Siemens Corporate Technology.
- ✧ Multi-Output Regularized Projection. Filed June 2005, Siemens Corporate Technology.
- ✧ Dirichlet Enhanced Latent Semantic Analysis. Filed October 2004, Siemens Corporate Technology.
- ✧ Vision-Based Document Segmentation. Filed July 2003, Microsoft Research Asia.

## AWARDS AND HONORS

- ✧ Student Travel Grant for KDD'2006
- ✧ Student Travel Grant for ICML'2006
- ✧ Best Paper Runner-Up & Best Student Paper Runner-Up Award in PKDD'2005
- ✧ Student Travel Grant for SIGIR'2005
- ✧ Student Travel Grant for ICML'2005
- ✧ Attendee of Machine Learning Summer School (MLSS'2004), Berder, France, Sep. 2004
- ✧ Student Travel Grant for SIGIR'2004

- ✧ Siemens Doctoral Scholarship, Sep. 2003 - Aug. 2006
- ✧ Outstanding Undergraduate Student of Beijing (<u><1%</u>), 2000
- ✧ Outstanding Undergraduate Student of Peking University (<u><2%</u>), 2000
- ✧ First Prize in National Mathematical Modeling Competition, 1999
- ✧ Award for Excellent Student of Peking University, in three successive years 1997 - 1999
- ✧ Scholarships, in three successive years 1997 - 1999
- ✧ Rank 3 (4 is the best) in National Computer Rank Examination Certificate, 1998
- ✧ First Prize in National Mathematics Competition for High School Students, 1995

## COMPUTER SKILLS

Languages mastered: C, C++, Java, MATLAB, .NET, Perl, Maple

Systems experienced: MS Windows, Linux

## OTHERS

- ✧ Fluent in Chinese and English. Read German and French
- ✧ Master accordion. Degree 5 (out of 10) in the National Accordion Certificate
- ✧ Enjoy badminton, table tennis, swimming, traveling and reading

## REFERENCES

Available upon request.