Raju Balakrishnan

CONTACT INFORMATION

Groupon
3101 Park Boulevard
Palo Alto, CA 94306 USA

Phone: (623)734-4794 (cell)
Fax: (650)433-4700 (office)
raju@groupon.com, raju.balakrishnan@gmail.com
http://rajubalakrishnan.com/

RESEARCH AREAS

[**Learning for Web Commerce, Big Data Integration and Ranking, Information Analysis on Networked Data, Auction Mechanisms for Online Ads, Real Time Bidding of Display Ads**] My research areas are emerging ranking and learning problems for the large scale semi-structured data. Specifically, I focus on analysis and ranking of the commercial data on the web, networked data analysis, and design and analysis of auction and bidding mechanisms of online ads.

PROFESSIONAL EXPERIENCE

Data Scientist, Groupon, Palo Alto, USA. Aug 2012-*till date.*

- Ranking and recommendation of daily deals.
- Deriving actionable inputs from merchants and consumer data.
- Developing machine-learned models for sales automation.

Software Engineer, Bangalore, IBM India. April 2004 - Dec 2005.

- Development of IBM DB2 database tools (DB/2 Information Management Group, Software Labs).
- Application oriented research in unstructured information management and information retrieval (Technology Incubation Center, Software Labs).
- System requirement analysis, design, and testing (System Engineer, Telecom and Media).

Summer Intern, Yahoo! Labs, Bangalore, India. June-Aug 2011.

Summer Intern, IBM Research, Delhi, India. June-Aug 2007.

EDUCATION

PhD, Arizona State University (Aug 2006—*July 2012*)
Department of Computer Science and Engineering
Tempe, Arizona, USA

- *Dissertation Topic:* Trust and Profit Sensitive Ranking for the Deep Web and Online Advertisements
- *Advisor: Prof.* Subbarao Kambhampati
- *Committee:* Prof. Subbarao Kambhampati (chair, ASU, *AAAI Fellow*), Dr Yi Chen (ASU), Prof. Huan Liu (ASU, *IEEE Fellow*), Dr AnHai Doan (University of Wisconsin, Madison)
- *CGPA:* 4.0/4.0
- *Year of Graduation:* August 2012

Bachelors of Engineering, Cochin University of Science and Technology (1997—2001)
College of Engineering, Model
Cochin, Kerala, India

- *Grade:* 76% (First Class with Distinction)
- *Year of Graduation:* August 2001

Research Awards

Best Poster Award, World Wide Web Conference (WWW'10), Raleigh, North Carolina, USA, 2010

Key Scientific Challenges Award, Yahoo! Research, Computational Advertisement Category, 2009-2010

First Inventor Award, by CEO, IBM 2006

Publications

Balakrishnan, R., and Kambhampati, S. *Click efficiency: a unified optimal ranking for online Ads and documents.* Journal of Intelligent Information Systems, Springer June 2015.

Balakrishnan, R., and Parekh R. *Learning to Predict Subject-Line Opens for Large-Scale Email Marketing.* (2014) In Proc of IEEE BigData 2014, Washington DC, USA.

Balakrishnan, R., and Bhatt R P. *Real-Time Bid Optimization for Group-Buying Ads.* (2014) ACM Transactions on Intelligent Systems and Technology, Special Issue on Computational Advertisement.

Ravikumar, S., Thalamadupula, K., **Balakrishnan, R.**, and Kambhampati, S. *RAProp: Ranking Tweets by Exploiting the Tweet/User/Web Ecosystem and Inter-Tweet Agreement.* (2013), In Proc. of CIKM 2013, San Francisco, USA.

Balakrishnan, R., Kambhampati, S, and Jha, M. *Assessing Relevance and Trust of the Deep Web Sources and Results Based on Inter-Source Agreement. .* (May 2013, Volume 7, Issue 2) ACM Transactions on the Web.

Balakrishnan, R., and Bhatt R P. *Real-Time Bid Optimization for Group-Buying Ads.* (2012) In Proc. of CIKM 2012, Maui Hawaii, USA.

Ravikumar, S., **Balakrishnan, R.**, and Kambhampati, S. *Ranking Tweets Considering Trust and Relevance.* (2012) Information Integration In the Web (2012), Scottsdale, USA.

Balakrishnan, R., and Kambhampati, S. *Factal: Integrating Deep Web Based on Trust and Relevance.* (2011) In Proc. of 20^{th} International World Wide Web Conference (WWW'11), Hyderabad, India.

Jha, M., **Balakrishnan, R.**, and Kambhampati, S. *Agreement Based Source Selection for the Multi-Domain Deep Web Integration* (2011) In Proceedings of COMAD'11, Bangalore, India (*Acceptance Rate=17%*).

Balakrishnan, R., and Kambhampati, S. *SourceRank: Trust and Relevance Based Source Selection for Deep Web.* (2010) In Proc. of 19^{th} International World Wide Web Conference (WWW'10), Pages 1055∼1056, Raleigh, North Carolina, USA *(Best Poster Award).*

Wolf G., Kalavagattu A., Khatri H., **Balakrishnan R.**, Chokshi B., Fan J., Chen Y. and Kambhampati S. *Query Rewriting Using Learned Data Dependencies* (2009). The International Journal of Very Large Databases (VLDB Journal) (Special issue on uncertain and probabilistic databases).

Balakrishnan, R., and Kambhampati, S. *Optimal Ad-Ranking for Profit Maximization.* (2008) In Proc. of 11th International Workshop on Web and Databases (WebDB'08). Vancouver, Canada (Co-Located with SIGMOD'08).

Nambiar U., Gupta H., **Balakrishnan R.**, Mohania M. *Helping Satisfy Multiple Objectives during a Service Desk Conversation.* In Proc. of ACM SIGMOD (2008), Vancouver, Canada.

Balakrishnan, R., Sahoo, R., *Lossless Compression for Large Scale Cluster Logs.* In Proc. of IEEE Parallel and Distributed Processing Symposium 2006, (IBM Research RC23902).

Balakrishnan, R., *Country wise classification of human names*, WSEAS International Conference on AI, Knowledge Engineering and Data Bases, 2006.

Balakrishnan, R., *Disulphide Bond Prediction in Proteins Based on Secondary Structure Constraints and Cysteine Separation*, CAINE-2005 (*US Patent 7672788*).

Balakrishnan, R., and Bhatt, R. P. *Real-Time Profit Maximization of Guaranteed Deals.* Yahoo Labs Technical Report, November 2011 (Under Review Conference Paper).

Balakrishnan, R., and Kambhampati, S. *Click Efficiency: A Unified Optimal Ranking for Online Ads and Documents.* arXiv Technical Report, July 2011 (Under Review Journal Article).

Balakrishnan, R., and Kambhampati, S. *On the Self-Similarity of Web Query Traffic: Evidence, Cause and Performance Implications.* Arizona State University Technical Report, 2009.

System Prototype

Factal, a deep web search engine based on our *SourceRank* algorithm.

Patents

Balakrishnan, R. *Disulphide Bond Connectivity in Proteins.* (US Patent: *7,672,788* Issue Date: Mar 2, 2010).

Balakrishnan, R, Bhatt R P. *System and Method for Dynamically Optimizing Profit for Guaranteed Deal Bidding.* (Patent Pending, Yahoo! Labs, USPTO: Y08119US00)

Balakrishnan, R, Parekh R G. *System and Method for Providing Optimized Leading Messages.* (Patent Pending, Groupon, Docket Number: 058407/437152)

Balakrishnan, R, Parekh R G, Zhang G, Tripathi V. *Method and System for Generating Purchase Recommendations Based on Purchase Category Associations.* (Patent Pending, Groupon, Docket Number: 0246APRV01US)

Gaurav G, Balakrishnan R, Parekh R G, Zhang G, *Apparatus and Method for Enhanced Message Targeting.* (Patent Pending, Groupon, Attorney Docket Number: 058407/458388)

Balakrishnan, R, Lai K, *et al.* *Method And Apparatus For Payment, Return On Investment, And Impact Reporting.* (Patent Pending, Groupon, Attorney Docket Number: 058407/436782)

Invited Talks

Microsoft Research. Ranking and Mechanism Design for Web Databases and On-line Advertisements, Bangalore, India, August 2011.

IBM Research. Trust and Profit Sensitive Ranking for Web Databases and On-line Advertisements, Bangalore, India, April 2011.

Yahoo! Research Graduate Student Summit. Ranking Deep Web Sources and Online Ads. Sunnyvale, CA, USA, September 2009.

IBM Research. Ranking Deep Web Sources and Online Ads, Delhi, India, April 2009.

Computer Skills

Languages: R, JAVA, Hive, C , C++, SQL, Pig Latin, PL/SQL, Perl, AWK, UML, Lisp.

Databases: Teradata, DB/2, Oracle, MySQL, DB2 Tools.

Web Technologies: J2EE, J2SE, Javascript, HTML, PHP, JSP.

Tools/Frameworks: Hadoop, Rational XDE, Latex, Eclipse IDE, MATLAB.

Web Servers: Tomcat, Websphere, JRUN.

Platforms: Windows, Linux, Mac.

Academic Experience

Graduate Courses: Information Retrieval, Machine Learning, Social Networking, Graph Theory (*Dept. of Mathematics*), Design of Engineering Experiments (*Dept. of Industrial Engg.*), Data on The Web, Database Internals, Game Theory, Combinatorial Algorithms and Intractability, Randomized Algorithms, Advanced Computer Networking, Computational Systems Biology.

Research Assistance: Fall 2008 to *till date*. Computer Science and Engineering, Arizona State University.

Teaching Assistance: Fall 2006 to Spring 2008. CSE 430: Operating Systems. Computer Science and Engineering, Arizona State University.

Bachelor of Technology in Computer Engineering: Model Engineering College, Cochin University of Science and Technology, Kerala, India. September 1997 - August 2001

- *Academic Projects:*
  1. Automatic Performance and Configuration Management of Local Area Networks (*Senior Level*)
  2. Small vocabulary Isolated Word Speech Recognition System (*Junior Level*)
- *Seminar:* Adaptive Resonance Theory
- *Significant Courses:* Computer Organization, Data Structures, Principles of Programming Languages, Operating Systems, Compiler Design, Algorithm Analysis and Design, Data Base Management Systems, Artificial Intelligence and Expert Systems, Software Engineering, Advanced Architecture and Parallel Processing, and Neural Networks.

Service

Referee: ACM Transactions on the Databases Systems, 2015.

Referee: ACM Transactions on The Web, 2013, 2014.

Referee: ACM Transactions on Intelligent Systems, 2013, 2014.

Senior Program Committee: International Joint Conference on Artificial Intelligence, 2013.

Reviewer: World Wide Web Conference, 2013.

Program Committee: Information Integration on the Web, 2012 (Co-Located with SIGMOD).

Significant Projects

Deal Relevance and Personalization for Groupon International.
Groupon, USA, Jan 2014-till date

Series of projects involving algorithm development and implementation for improved ranking and personalization of deals. Algorithms are based on collaborative filtering on purchase/click history of consumers.

Outcome: Number of deployed algorithms for improved revenue and consumer satisfaction.

Predicting Performance of email Subject Lines.
Groupon, USA, June 2013-Sept 2013

Predicting open rates of email-subject lines by learning from historical data.

Outcome: A random forest based model is developed and validated.

Predicting purchase-propensity of consumers for email-campaigns.
Groupon, USA, July 2013-current

Predictive models for customers with highest incremental probability to purchase with discounts to improve the revenue lift in email campaigns.

Outcome: Provided gradient boosted regression based models for prediction to CRM team.

Ranking Merchants for Improved Deal Closing.
Groupon, USA, Sept 2012-Dec 2012

The project was to improve models for predicting merchants probability to run daily deals, based on the publicly and privately available information about the merchants.

Outcome: The support vector machine based model performed better, and deployed in production.

Predicting Deal Performance.
Groupon, USA, Nov 2012-Dec 2012

The project was to build models for predicting deal performance of merchants, based on the merchant data, and past deals.

Outcome: Logistic regression based models were successful and deployed in production.

Real-Time Bid Optimization for Group-Buying Ads.
Yahoo Labs, Bangalore, Summer 2011

The optimal bid values for group buying ads requiring guarantees of minimum number of conversion within a fixed time period was unknown. The research project involved deriving expected profits, deriving optimal bids, and approximations to meet the stringent time constraints in real-time bidding

Outcome: Published paper in CIKM'12 and Patent Pending.

Real-time Profit Maximization for Call Centers
IBM India Research Lab, Summer 2007

Call-center agents face the difficult task of asking right questions to maximize the profits while minimizing the call duration. The system suggests questions to the agent at real time to satisfy these objectives. The suggestions are based on the noisy call transcripts generated by a speech recognition system and the inventory. The questions are formulated to maximize the expected profit and the information gain.

Outcome: Published paper in SIGMOD'08.

DB/2 Tools Development
IBM India Software Lab, Feb 2006-Aug 2006.

Development and Support for Various DB/2 Tools like Task Center, Journal, JDBC and ODBC Drivers, and Control Center. I was responsible for development of the above mentioned components.

Technology Incubation Center
IBM India Software Labs, April 2004 - Dec 2005.

Development of prototypes and prof of concept developments in collaboration with research and product groups. The group was working as an interface between the research and product groups in India. The significant projects were:

1. Unstructured Information Management Architecture (UIMA is the principal framework for Watson's Jeopardy!) based Language Identifier for Indian Languages.
2. UIMA based multi-modal Search Engine.
3. Log Monitor for Bluegene/L (the fastest super-compute at that time).

Outcome: IBM Research Report on Log Management for large clusters, two publications, and one US Patent (pending).

Telecom and Media

IBM Services India, August 2001-April 2004.

Development and maintenance of data services contracting automation for AT&T. I performed various roles including system engineer, system tester, developer, and tester in various projects. The significant projects are:

1. Electronic Data Service Offers
   Role: System Engineer.
   Skills:Java, UNIX, SQL, UML, Rational Rose, ORACLE, XML.
   Project added the capability for the system to handle High Speed Packet Services and Private Line services with the capability of automatic biller upload.
2. eCRM Call For MCN Validation
   Role: Designer, Developer.
   Skills: Java, UNIX, SQL.
   An enhancement project involving addition of XML, Connect Direct and JDBC interfaces to add the capability to handle the Customer Validation Requests send by eCRM platform.
3. Electronic Customer Relationship Management
   Role: Designer, Developer.
   Skills: Java, UNIX, UML, SQL.
   This project was designed to add new interface with newly introduces eCRM platform for contracting. An enhancement project involving addition of RMI/Connect Direct, XML and JDBC interfaces to add the new workflow to handle the deals send by eCRM platform.
4. GCO Admin Enhancements
   Role: Developer.
   Skills: UNIX, SQL.
   The was to increase the performance and maintainability of Generic Contract Tariff module tenfold i.e. to decrease response time to 10 of the original. This involved redesign of the entire module with very ambitions performance objectives.
5. Contracting Rule Automation
   Role: Developer.
   Skills: Java, UNIX, SQL.
   An enhancement project involving addition of XML, Connect Direct and JDBC interfaces to add the capability to handle the MCN Validation Requests send by eCRM platform.

CITIZENSHIP Indian Citizenship. H1B, USA.