Praveen Ravichandran

curriculum vitae

30 Regent St, Apt 316
Jersey City, NJ - 07302
☎ +1 (201) 314 6682
✉ pcr@udel.edu
🖳 pchandar.github.io

I am interested in the field of information retrieval, natural language processing and machine learning. More specifically, I am interested in the study and development of effective and efficient evaluation techniques that help measure how well retrieval systems satisfy user's information needs. I have worked on research problems including evaluation of retrieval and question answering systems, web search, recommendation systems, and dialogue system for healthcare and scheduling. As a result of this work, I have created open-source tools, contributed to commercial software and published in leading academic conferences.

My dissertation combines insights from information retrieval, user modeling, and statistics to improve the effectiveness of search evaluation, leading to a more accurate measurement of system effectiveness and thereby revolutionizing the way people seek information.

Education

Aug 2014   **Ph.D., Computer Science**, *University of Delaware*, Newark, DE.
Thesis: Novelty and Diversity in Search Results
Advisor: Dr. Benjamin A Carterette

May 2008   **Bachelor of Engineering in Computer Science**, *Anna University*, Chennai, India.
First Class with Distinction

Professional Experience

Apr '15 – Present   **Data Scientist**, *X.AI Inc.*, New York, NY.
Worked on various NLU components of a dailouge system; helped build an NLP pipeline to extract named entities; hepled hire new talent.

Sept '14 – Apr'15   **Postdoctoral Research Scientist**, *Columbia University*, New York, NY.
Built a framework to analyze and categorize informational needs of clincal researchers; developed algorithms to semi-automatically enrich medical terminologies.

Sept '13 – May '14   **Teaching Assistant**, *University of Delaware*, Newark, DE.

Feb '09 – May '14   **Research Assistant**, *University of Delaware*, Newark, DE.
Worked on research problems relating to novelty and diversity in search results; conducted user studies to better understand user needs; developed a preference framework for novelty evaluation.

Summer 2013   **Research Intern**, *IBM T.J. Watson Research Center*, Yorktown Heights, NY.
Worked on ways to evaluate answer snippets clustering. Focused on developing annotation strategies to build gold standards against which the algorithms could be evaluated.

Summer 2010   **Search Relevance Intern**, *OneRiot Inc*, Palo Alto, CA.
Worked on ways to improve the relevance of OneRiot's real time search. Primary focus was on evaluation, in both the context of real-time search and ad results.

Research

Research Interests

Information Retrieval, User Modeling, Natural Language Processing, Machine Learning

Peer-Reviewed Publications

JBI '16   **DREAM: Classification Scheme for Dialog Acts in Clinical Research Query Mediation** Hoxha J, Chandar P, He Z, J. Cimino, D. Hanauer, Weng C. (2016) Vol-59 Issue-1 pp. 89–101)

AMIA '15   **Simulation-based Evaluation of the Generalizability Index for Study Traits** Z. He, P. Chandar, P. Ryan, C. Weng. *Proceedings of AMIA 2015 Annual Symposium (2015), pp.593–602*

AMIA '15   **Similarity-based Recommendation of New Concepts to a Terminology** P. Chandar, A. Yaman, J. Hoxha, Z. He, C. Weng. *Proceedings of AMIA 2015 Annual Symposium (2015), pp.386–395*

AMIA '15   **What Are Frequent Data Requests from Researchers? A Conceptual Model of Researchers' EHR Data Needs for Comparative Effectiveness Research Podium abstract**G. Hruby, P. Chandar, J. Hoxha, E. Mandonca, D. Hanauer, C. Weng. *Proceedings of AMIA 2015 Annual Symposium (2015), Podium Abstract*

SIGIR '15   **Document Comprehensiveness and User Preferences in Novelty Search Tasks** A. Bah, P. Chandar, B. Carterette. *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval (2015), pp. 735–738*

SIGIR '13   **Preference Based Evaluation Measures for Novelty and Diversity** P. Chandar, B. Carterette. In *Proceedings of the 36th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. (2013), pp. 413–422.*

SIGIR '13   **Document Features Predicting Assessor Disagreement** P. Chandar, W. Webber, B. Carterette. In *Proceedings of the 36th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (2013), pp. 745–748.*

CIKM '12   **Alternative Assessor Disagreement and Retrieval Depth** W. Webber, P. Chandar, B. Carterette. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management (2012), pp. 125–134.*

SIGIR '12   **Using Preference Judgments for Novel Document Retrieval** P. Chandar, B. Carterette. In *Proceedings of the 35th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (2012), pp. 861–870.*

SIGIR '12   **Using PageRank to Infer User Preferences** P. Chandar, B. Carterette. In *Proceedings of the 35th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (2012), pp. 1167–1168.*

workshop @ WSDM '12   **What Qualities Do Users Prefer in Diversity Rankings?** P. Chandar, B. Carterette. *Workshop on Diversity in Document Retrieval at the fifth ACM WSDM Conference (2012).*

workshop @ ECIR '11   **Analysis of Various Evaluation Measures for Diversity** P. Chandar, B. Carterette. *Workshop on Diversity in Document Retrieval at 33rd European Conference on Information Retrieval (ECIR) (2011).*

SIGIR '10   **Diversification of Search Results using Webgraphs** P. Chandar, B. Carterette. In *Proceedings of the 33rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (2010), pp. 869–870.*

CIKM '09   **Probabilistic Models for Facet Retrieval** B. Carterette, P. Chandar. In *Proceedings of the 18th ACM International Conference on Information and Knowledge Management 2009, pp. 1287–1296.*

Technical Reports and Other Publications

NTCIR '14   **Udel @ NTCIR-11 IMine Track** A. Bah, P. Chandar, B. Carterette. *Proceedings of the 11th NTCIR Conference (2014).*

TREC '13   **University of Delaware at TREC 2013** A. Bah, P. Chandar, K. Sabhnani, M. Zengin, B. Carterette. *Proceedings of the 22nd Text Retrieval Conference (2013).*

TREC '12   **University of Delaware at TREC 2012** A. Bah, P. Chandar, N. Kumar, A. Rao, D. Zhu, B. Carterette. *Proceedings of the 21st Text Retrieval Conference (2012).*

TREC '11   **Implicit Feedback and Document Filtering for IR Over Query Sessions** B. Carterette, P. Chandar. *Proceedings of the 20th Text Retrieval Conference (2011).*

TREC '10   **Sessions, Diversity, and Ad Hoc Retrieval** B. Carterette, P. Chandar. *Proceedings of the 19th Text Retrieval Conference (2010).*

TREC '09   **Ad Hoc and Diversity Retrieval at the University of Delaware** P. Chandar, A. Kailasam, D. Muppaneni, L. Thota, B. Carterette. *Proceedings of the 18th Text Retrieval Conference (2009).*

TREC '09   **Minimal Test Collections for Relevance Feedback**. B. Carterette, P. Chandar, A. Kailasam, D. Muppaneni, L. Thota. *Proceedings of the 18th Text Retrieval Conference (2009).*

Research Projects

Models and Measures for Novel And Diverse Search results
Traditional models of information retrieval assume documents are independently relevant. But modeling documents as independently relevant is more likely to produce a ranking with a high degree of redundancy; the amount of novel information available to the user may be minimal as they traverse down a ranked list. This research project attempts to remedy this with new models of document interdependence and new evaluation measures. There are three threads running through this work: (1) the models of diversity, novelty, and redundancy that will be needed to implement ranking algorithms; (2) measurements of diversity, novelty, and redundancy in a ranking of documents; and (3) optimizing model structures and parameters to the measures. As part of this project, we developed statistical models to reduce the degree of redundancy in a rank list, analyzed various evaluation frameworks, designed and conducted user studies to better understand user's need and proposed a novel preference based evaluation framework. This project resulted in various conference publications and a thesis.

Predicting Assessor Disagreement in Information Retrieval
Assessors are well known to disagree frequently on the relevance of documents to a topic, but the factors leading to assessor disagreement are still poorly understood. As part of this project, we studied the relationship between assessor disagreement and various factors such as readability, cohesiveness, and rank at which a document is returned by a set of retrieval systems (meta-rank). To this end, we proposed a logistic regression predictive model of second assessor disagreement given meta-rank and initially-assessed relevance. This project is resulted in various conference publications.

Evaluation of Answer Snippet Clusters

Intelligence analysts are often faced by scenarios where they are confronted by informal sources of text including blogs and fourms, and would like to study relationships among people involved in the discussion, points of view expressed regarding a specific event. The problem involves a deep analysis of the large amounts of text. A typical system build for this problem returns a list of relevant answer snippets for a given natural language question and it is often a good idea to cluster (group) the returned answer snippets based on some criteria. In this project, our focus was on developing ways to evaluate the snippet cluster retured by the system. This project resulted in a novel user interface to annotate snippet clusters.

Teaching

Teaching Assistant, University of Delaware

CISC 106   **Introductory Computer Science for Engineers**

- Instructor: Debra Yarrington
- Assisted in creating lab and homework assignments.
- Conducted lab sessions and graded assignments.

CISC 672/471   **Compiler Design**

- Instructor: Prof. Lori Pollock
- Lectured on the following topics: Bottom-up parsing, and Abstract syntax trees
- Assisted in construction of lab projects and homework assignments.
- Held office hours, responding to student concerns on course material

Presentations and Talks

- SIG-AI, (Special Interest Group on Artificial Intelligence, weekly UD seminar), 2013, University of Delaware. Annotation Strategies for Clustering Answer Snippets
- SIGIR, 2012, Portland, USA. Using Preference Judgments for Novel Document Retrieval.
- WSDM, 2012, Seattle, USA. What Qualities Do Users Prefer in Diversity Rankings?
- SIG-AI, (Special Interest Group on Artificial Intelligence, weekly UD seminar), 2012, University of Delaware. Evaluating Ad-hoc and Diversity Using Preference Judgments.
- ECIR, 2011, Dublin, Ireland. Analysis of Various Evaluation Measures for Diversity
- SIG-NLP, (Special Interest Group on Natural Language Processing, weekly UD seminar), 2011, University of Delaware. Evaluation Frameworks in Novelty and Diversity.
- TREC, 2009, Gaithersburg, Maryland, USA. Ad-hoc and Diversity Retrieval.

Proficiencies

Programming   Java, Scala, R, Python, C++, C, Perl, Lisp
Search Tools   Indri/Lemur, Terrier, Lucene
NLP/ML Tools   StanfordNLP, Factorie, scikit-learn

Service and Awards

Committees and Reviewing

- Reviewer, IEEE Transactions on Knowledge and Data Engineering

- Reviewer, Information Sciences
- Program Committee Member, Asia Information Retrieval Societies Conference (AIRS), 2013
- Program Committee Member, Conference on Information and Knowledge Management (CIKM), 2014
- Program Committee Member, Conference on Information and Knowledge Management (CIKM), 2015

Travel Grants

- **Professional Development Award, [2010]**
  Competitive monetary award given to graduate students for financial assistance to present and attend a major conference in their field
- **SIGIR Travel Grant, [2010, 2012]**
  Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in 2010 and 2012

References

- Available upon request