

Curriculum Vitae

Grace, Hui Yang

Telephone: (+1)412-268-8492

Email: huiyang@cs.cmu.edu

Language Technologies Institute
School of Computer Science
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA, USA, 15213

Goal

I would like to be a top researcher and educator for the advance of computer technology and its influence on our society.

Education

- Ph.D Candidate at School of Computer Science, Carnegie Mellon University, 2004 – now (expected 2009)
- Master of Science, School of Computing, National University of Singapore, 2001-2003
- Bachelor of Science with Honors (First Class Honors), School of Computer Science, National University of Singapore, 2001
- Bachelor of Science, School of Computer Science, National University of Singapore, 1997-2000

Honors

- Carnegie Mellon University LTI Scholarship (2004- now)
- First Class Honors in Computer Science, National University of Singapore (2001)
- Distinguished Honors List, National University of Singapore (2000)
- Singapore Ministry of Education Scholarship (1996-2001)
- Excellent Youth Award in Province, China (1995)
- 1st Prize in Province, National Mathematics Olympics Competition, China (1993)
- 2st Prize in Province, National Chemistry Olympics Competition, China (1993)
- 1st Prize in Province, Essay Competition, China (1992)

Research Interests

Information Retrieval, Text Mining, Machine Learning, Natural Language Processing, Question Answering

Research Experience

- Ontology Learning, Carnegie Mellon University (Aug 2006-now)

Ontology learning is an emerging hot topic in the intersection of research areas of Semantic Web, Knowledge Acquisition, Knowledge Management, Information Retrieval, Natural Language Processing and Machine Learning. From raw text document collections, automatic or interactive ontology acquisition and construction are done through a general framework embedding techniques from statistical learning, linguistic analysis, and multiple resources like the Web and WordNet.

The work has been published in DG.O 2008.

- Sentiment Detection and Opinion Detection, Carnegie Mellon University (May 2006-Aug 2006)

Sentiment Detection and Opinion Detection in product review, movie review, blogspace and political comments is a challenging task. By Bayesian Logistic Regression, the system classifies whether a document contains an opinion and the sentiments of the opinion. Linguistic features and statistical language features are combined to improve the detection accuracy.

The work has participated in TREC 2006 evaluation and published in TREC 2006 processing.

- Blog Retrieval, Carnegie Mellon University (May 2006- Aug 2006)

Blogspace is an emerging place that people write their daily thoughts and experiences. It connects the people together and will greatly influence both the social and technical aspects of our life. Blog retrieval and information seeking is hence very important. By exploring the linkage relationship, RSS feeds and the content structure, a more flexible framework is developed to enhance the famous Indri search engine.

- Near-Duplicate Detection in eRulmaking, Carnegie Mellon University (2004-2007)

The U.S. regulatory agencies are required to read and solicit every single piece of public comments to the proposed rules. To save the human effort in the rulemaking process, near-duplicate detection is developed via a semi-supervised clustering approach, which allows flexibly incorporating constraints into the clustering process to achieve a better clustering accuracy.

The work has been reported in Digital News Journal (2004 Aug) and has been published in SIGIR 2006, DG.O 2006, and DG.O 2005.

- Multimedia Information Retrieval, Carnegie Mellon University (2004)

News Video collection contains thousands of hours of videos, which is a combination of text scripts, audio, image, and video sequences. To find a qualified video sequences matching with the use query, the system applies text analysis, audio analysis, speech recognition and image processing. A comparison of unimodal, multi-modal and multi-concept classifiers of feature extraction is studied. Both visual-only features and multi-modal video features are also explored in the search process.

The work has participated in the TRECVID 2004 evaluation and won the second place. It has been published in TRECVID 2004.

- Question Answering, National University of Singapore (2001-2004)
TREC Question Answering track is a task that answering factoid, definition, list questions created by NIST assessors. By taking the event-based question answering approach, the system makes use of external recourses such as WordNet definitions, synonyms and web snippets to gather the additional knowledge about the question-answer event. It then finds answers by question-answer type-matching, linguistic chaining.

The system participated in TREC2002, TREC2003, TREC2004 evaluations and consistently won the 2nd place in TREC QA evaluation for 3 years.

The work has been published in SIGIR 2003, TREC 2002, TREC 2003, COLING 2004, EACL 2003, WWW 2003, SIGIR 2004.

- VideoQA: Question Answering on News Video, National University of Singapore (2003-2004)
The work research explores the use of question answering (QA) techniques to support personalized news video retrieval. The system uses short natural language questions with implicit constraints on contents, context, duration, and genre of expected videos. It returns short precise news video summaries as answers. It uses multi-modal features, including visual, audio, textual, and external resources, to help correct speech recognition errors and to perform precise question answering.

The work has been published in ACM MM 2003 conference.

- Multimedia Information Retrieval, National University of Singapore (2003-2004)
The work explores the application of text information retrieval techniques to the multimedia domain. It analyses the text query issued by the users and extracts relevant video stories based on ASR, and external resources like WordNet and related news articles on the web. It then acts as a concept filter, which eliminates the irrelevant video shots in the stories retrieved by text query system. The system finally re-ranks the retrieved shots using the image and video retrieval techniques with relevance feedback.

The work has been published in TRECVID 2003.

- Online Streaming Video Broadcasting and Recording, (2000-2001)
The work sets up an online video station by capturing the analogue video signals broadcasted from local television stations. It converts the analogue signals into digital signals and allows the user to view and select from different stations. A video recording feature also highlights the capability of the system being a ready-to-be-commercialized product. Main research effort was put in the Speech and video synchronization.

The work was a final year undergraduate individual research project and won full marks from the project evaluation.

Teaching Experience

Information Retrieval, Carnegie Mellon University, Spring 2008

Multimedia Technologies, National University of Singapore, 2001-2002

Artificial Intelligence, National University of Singapore, 2003-2004

Software Engineering, National University of Singapore, 2001-2004

Programming in Java, National University of Singapore, 2000-2001

(All courses are conducted in English)

Working Experience

Research Assistant (Aug 2004-now), Carnegie Mellon University

Research Assistant (July 2004-Aug 2004), National University of Singapore

Teaching Assistant (July 2001-July 2004), National University of Singapore

System Engineer (May 2001-July 2001), National University of Singapore

Part-time System Engineer (May 2001- July 2001), GlobalID Asia

Part-time Teaching Assistant (July 2000- May 2001), National University of Singapore

System Analyst (May 1999 – Jan 2000), Internship at Singapore Telecommunication

Publication

- Hui Yang and Jamie Callan. "Ontology generation for large email collections." Proceedings of the Eighth National Conference on Digital Government Research. Montreal. May, 2008.
- Hui Yang , Luo Si, Jamie Callan, "Knowledge Transfer and Opinion Detection in the TREC2006 Blog Track", In the Notebook of Text REtrieval Conference 2006 (TREC2006), Gathersburgh, MD, Nov 14-17 2006.
- Hui Yang , Jamie Callan, "Near-Duplicate Detection by Instance-level Constrained Clustering", In Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR2006), Seattle, WA, Aug 6-11 2006.
- Hui Yang , Jamie Callan, Stuart Shulman, "Next Steps in Near-Duplicate Detection for eRulemaking", In Proceedings of the 6th National Conference on Digital Government Research (DG.O2006), San Diego, California , May 21-24 2006.
- Hui Yang, Jamie Callan, Stuart Shulman. "DURIAN: A demo for near-duplicate detection" (demo description). Proceedings of the Sixth National Conference on Digital Government Research (DG.O2006). Atlanta, GA. 2006.
- Hui Yang , Jamie Callan, "Near-Duplicate Detection for eRulemaking", In Proceedings of the 5th National Conference on Digital Government Research (DG.O2005), Atlanta, GA, USA, 15-18 May 2005.
- Alex. Hauptmann, Ming-Yu Chen, Mike Christel, C. Huang, W.-H. Lin, Toubin Ng, N. Papernick, A. Velivelli, Jun Yang, Rong Yan, Hui Yang, and H.D. Wactlar, "Confounded Expectations: Informedia at TRECVID 2004", In the notebook of the 13th Text REtrieval Conference Video Workshop (TRECVID2004), Maryland, USA, 15-16 Nov 2004.

- Hui Yang, Lekha Chaison, Yunlong Zhao, Shi-Yong Neo, Tat-Seng Chua, "VideoQA: Question Answering on News Video", In the Proceedings of the Eleventh Annual ACM International Conference on Multimedia (ACMM'2003), Berkeley, California, USA, 2-8 Nov 2003.
- Tat-Seng Chua, Yunlong Zhao, Lekha Chaisorn, Chun-Keat Koh, Hui Yang, Huaxin Xu, Qi Tian, "TREC 2003 Video Retrieval and Story Segmentation Task at NUS PRIS", In the notebook of the 12th Text REtrieval Conference Video Workshop (TRECVID'2003), Maryland, USA.
- Hui Yang, Tat-Seng Chua, "Web-based List Question Answering", In the Proceedings of the 20th International Conference on Computational Linguistics (Coling'2004), Geneva, Switzerland, 23-27 Aug 2004.
- Hui Yang, Tat-Seng Chua, "Effective Web Page Classification for Finding List Answers", In the Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'2004), Sheffield, UK, 25-29 Aug 2004.
- Hui Yang, Tat-Seng Chua, "FAD on the Web: Find All Distinct Answers", In the Proceedings of the Thirteenth International World Wide Web Conference (WWW'2004), New York, USA, 17-22 May 2004.
- Hui Yang , Tat-Seng Chua, Shuguang Wang, Chun-Keat Koh, "Structured Use of External Knowledge for Event-based Open Domain Question Answering", In the Proceedings of the Twenty-Sixth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'2003), Toronto, CANADA, 28 July-1 Aug 2003.
- Hui Yang , Tat-Seng Chua, Shuguang Wang, "Modeling Web Knowledge for Answering Event-based Questions", In the Proceedings of the Twelfth International World Wide Web Conference (WWW'2003), Budapest, HUNGARY, 20-24 May 2003.
- Hui Yang , Tat-Seng Chua, "QUALIFIER: Question Answering by Lexical Fabric and External Resources", In the Proceedings of the Tenth Conference of the European Chapter of the Association for Computational Linguistics (EACL'2003), Budapest, HUNGARY, 12-17 April 2003. page 363-370.
- Hui Yang, Hang Cui, Min-Yen Kan, Mstislav Maslennikov, Long Qiu, Tat-Seng Chua, "QUALIFIER in TREC-12 QA Main Task", In the notebook of the 12th Text REtrieval Conference (TREC'2003), Maryland, USA.
- Hui Yang , Tat-Seng Chua, "The Integration of Lexical Knowledge and External Resources for Question Answering", In the Proceedings of the Eleventh Text REtrieval Conference (TREC'2002), Maryland, USA, 19-22 Nov 2002, page 155-161.

Professional Services

- Reviewer of ACM Special Interest Group on Information Retrieval (SIGIR), 2008
- Reviewer of ACM Special Interest Group on Information Retrieval (SIGIR), 2004

Programming Skills

C++, C, Java, Perl, Matlab, Microsoft C++, Microsoft Basic, PHP, Javascript, HTML

Language Skills

English (fluent), Mandarin Chinese (native), Cantonese (fair), Japanese (fair)

Extracurricular Activities

- Vocal in SCS Day Talent Show 2006, performing Chinese Traditional Huangmei Opera, Carnegie Mellon University
- Skydiving committee member, Carnegie Mellon University
- Skiing committee member, Carnegie Mellon University
- Explorer committee member, Carnegie Mellon University
- Chair of Song-writing group in Kent Ridge Hall, National University of Singapore
- Song-writer and Vocal in Kent Ridge Hall Inspire, National University of Singapore
- Vocal in University Song-writing group, National University of Singapore
- Vocal in Performance in Takashimaya Shopping Center, Singapore, 1997-1998
- Actress and Script writer in Kent Ridge Hall Drama group, National University of Singapore
- Cheerleader in cheerleading group in Kent Ridge Hall, National University of Singapore
- Designer and Cloth Maker in Kent Ridge Hall Production Wardrobe Committee, National University of Singapore
- Captain of Female Soccer Team in Kent Ridge Hall Block E, National University of Singapore
- Member of Tag-of-war Team in Kent Ridge Hall, National University of Singapore