Wencan Luo (wencan@cs.pitt.edu, +1412-973-1429)

Department of Computer Science, University of Pittsburgh

210 South Bouquet St. Room 5420, Pittsburgh, PA, 15260, USA

Research Interests:

Natural Language Processing, Machine Learning, Automatic Summarization

Education Background

PhD: Sep 2011 – now (6th year)   GPA: 3.96 (4.0)

Department of Computer Science at **University of Pittsburgh**

Advisor: [Diane Litman](#)

Master: Sep 2008 – Jun 2011

Institute of Software at **University of Chinese Academy of Sciences**

Advisor: [Feng Tian](#) (Full Professor)

Bachelor: Sep 2004 - Jun 2008   Ranked: 5/157   GPA: 3.67 (4.0)

Computer Science and Technology at **University of Science & Technology Beijing**

Thesis Topic: Automatic Robot based on Machine Vision (excellent thesis)

Advisor: Zhiming Wang (Associate Professor)

Internships

Jun 2015 - Aug 2015

Organization: Nuance Communications (U.S.)

Topic: Spoken Language Understanding

Referee: Fang Yang (Fan.Yang@nuance.com)

May 2013 - Aug 2013

Organization: Bosch Research and Technology Center (U.S.)

Topic: Natural Language Understanding (NLU)

Referee: Lin Zhao (lin.zhao@us.bosch.com)

Feb 2009 - Aug 2009

Organization: France Telecom R&D Beijing

Topic: Natural Language Processing

Referee: XinNian Mao (maoliwell@vip.sina.com)

Project & Research Experiences

Thesis: Automatic Summarization for Student Course Responses

Role: Researcher, advised by [Diane Litman](#)   **Date:** Jan 2014 – Present

Project Outline: We built a mobile application (CourseMIRROR) to automatically collect and summarize student responses. We proposed different approaches to enhance student response summarization at multiple levels of granularity. At the

sentence level, we propose a novel summarization algorithm by extending ILP-based framework with matrix imputation to address the challenge of lexical variety among student responses. At the phrase level, we propose a phrase summarization algorithm by a combination of phrase extraction, phrase clustering and phrase ranking in order to meet the need of aggregating and displaying student responses in a mobile application. The summarization is created from phrases rather than sentences and consists of three key steps. Last, we propose a quantitative phrase summarization algorithm in order to estimate the number of students who semantically mention the phrases in a summary.

Natural language question answering engine with DBpedia (Google Summer Code 2014)

Role: Researcher, advised by Axel Ngomo   **Date:** May 2014 – Aug 2014

Project Outline: In this project, we have built a Chinese question answering (QA) engine for DBpedia. The basic idea is to convert a natural language question into a structured SPARQL query and then send the query to the structured data in DBpedia. There are two major challenges we have addressed. The first one is to disambiguating named entities and link them to DBpedia. To solve it, a graph-based disambiguation model is proposed, which is language-independent and combines breadth-first search and the well-known HITS algorithm. The code is available at https://github.com/wencanluo/AGDISTIS.   The second challenge is generating SPARQL queries. To address it, we leverage human-derived rules based on part-of-speech tagging to convert questions to semantic representations and generate queries based on automatic built templates. The complete QA engine is available at https://github.com/wencanluo/AutoSPARQL.

e-Chimera: Rapid In-Situ Mobile Experimentation through Integrated Design, Test, and Deployment

Role: Researcher, advised by Jingtao Wang   **Date:** Jan 2012 – Dec 2013

Project Outline: The advent of smartphones and Internet tablets provides unique opportunities for running large-scale, in-situ experiments in the wild. However, the steep learning curve of building mobile applications makes it difficult for researchers in most fields to design and deploy cognitive, behavioral, and social experiments on mobile platforms. Inspired by contextual inquiries with domain researchers and a comprehensive literature survey, we present e-Chimera, a visual end-user programming environment for designing and prototyping experiments on mobile devices. The e-Chimera IDE provides a statechart-based, multi-level visual language to create experiments at the workflow, trial, and screen interaction level. The e-Chimera runtime, driven by a Domain-Specific Language (DSL), enables high fidelity, cross-platform application deployment. Overall, e-Chimera integrates design, test, deployment, and analysis of mobile experiments.
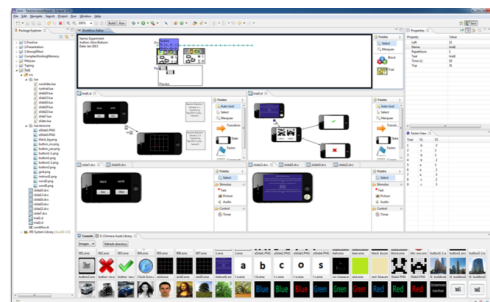
Reducing Annotation Effort on Unbalanced Corpus based on Cost Matrix

Role: Researcher, advised by Diane Litman   **Date:** Jan 2012 – May 2012

Project Outline: Annotated corpora play a significant role in many NLP applications. However, annotation by humans is time-consuming and costly. In this project, a high recall predictor based on a cost sensitive learner is proposed as a method to semi-automate the annotation of unbalanced classes such as uncertainty. We demonstrate the effectiveness of our approach in the context of one form of unbalanced task: annotation of transcribed human-human dialogues for presence/absence of uncertainty. In two data sets, our cost-matrix based method of uncertainty annotation achieved high levels of recall while maintaining acceptable levels of accuracy. The method is able to reduce human annotation effort by about 80% without a significant loss in data quality, as demonstrated by an extrinsic evaluation showing that results originally achieved using manually-obtained uncertainty annotations can be replicated using semi-automatically obtained uncertainty annotations.

Multi-label Classification Implementation (A Project for Google Summer Code 2011)
Role: Developer, advised by [Matija Polajnar](#)   **Date:** Jun 2011 – Aug 2011
Core Algorithm: Multi-label Classification (Binary Relevance, BR-KNN, etc.)
Project Outline: It's one of the projects of Google Summer Code 2011 for [Orange](#). The main goal is to extend the Orange to support multi-label, from dataset support to algorithms, from testing to evaluation, from GUI widgets and documentation. During these three months, two transformation methods and two algorithm adaptation methods for multi-label classification are implemented. Along with the methods, their widgets are also added. As the evaluation metrics for multi-label data are different from the single-label ones, new evaluation measures are supported. The codes with documents are available in [SVN branch](#).

Badminton Training System based on Physiological Computing
Role: Team leader, advised by [Feng Tian](#)   **Date:** Sep 2009 – Jun 2011
Core Algorithm: Signal Processing, Machine Learning (C4.5)
Project Outline: Firstly, a simple system is implemented based on accelerating sensors, which can recognize four kinds of activities-sit, stand, walk, and run. Furthermore, a badminton training assistant leveraging psychophysical data in monitoring and analyzing the states of Badminton Players is developed, named Psychophysical Plan. It can collect ECG, EMG, Bioimpedance, 3-axis Accelerations of the player, provide feedback of the data to coaches, and help them monitor and analyze the physiological states of the player.

Mobile Learning Approaches via Culturally Inspired Group Games
Responsibility: Developer   **Date:** Jun 2009 – Sep 2009
Core Algorithm: $1, stroke order recognition
Project Outline: Inspired by mobile group learning games, two mobile phone games - Multimedia Word and Drumming Stroke, are developed, and are aiming to help children to learn Chinese characters including the right stroke order. These two games are implemented in Nokia N800, which are written in C using GTK/GNOME, GStreamer. Evaluations reveal that the two games have the potential to enhance the intuitiveness and engagement of traditional games, and at the same time, may significantly improve children's literacy acquisition.

ABU Asia-Pacific Robot Contest 2008

Responsibility: Developer   **Date:** Jul 2007 – Aug 2008

Core Algorithm: A*, Computer Vision

Project Outline: In this senior year, I joined the team for the Asia-Pacific Robot Contest. We built the robot from scratch, and I wrote codes giving them the capability of task schedule, self-localization, path planning, trajectory tracking, as well as vision based object detection and obstacle avoidance. The most impressive two contributions I made, one is the algorithm for curved path planning. Another is the obstacle avoidance based on vision. At last, we qualified the 8th-finals.

Natural Language Processing Toolkit (A Project in France Telecom)

Responsibility: Developer   **Date:** Feb 2009 – Jul 2009

Core Algorithm: HMM, Double Array, CRF

Project Outline: The NLP toolkit I developed in this half year includes the following modules:
1) N-gram language model trained with SRILM toolkit. 2) A Part-of-Speech tagger based on the language model. The HMM and Viterbi decode algorithm are used in the tagger and the accuracy is 97.47%. 3) A Chinese Word Segmentation with entity-name recognition using HMM. 4) Convert Chinese characters to Pinyin and Pinyin to characters. 5) Simplified and Traditional Chinese transformation. 6) Spell Correction using Edit-Distance and Double Array. 7) Entity Recognition, including Chinese Name, Place Name and Organization name.

Awards and Achievements

2016	Best Student Paper Award Nominee at the 29th Flairs Conference Andrew Mellon Fellowship for Academic Year 2016-2017
2014	Accomplishment of Google Summer Code 2014
2012	Arts & Sciences Graduate Fellowship
2011	Outstanding Graduate of Graduate University of the Chinese Academy of Sciences (3%) Accomplishment of Google Summer Code 2011
2010	Excellent Student Honor of Graduate University of the Chinese Academy of Sciences (15%)
2008	Beijing Area College Outstanding Graduate
2007	People's Top Grade Scholarship
2006	Scholarship of China Excellent Students Awards
2005	People's Top Grade Scholarship The third prize at Beijing 16th Undergraduate Student Mathematics Competition Excellent Students Awards The third prize at the 16th Beijing College Students Mathematics Olympiad Excellent Student in Summer Social Practice
2004	Scholarship for Excellent Freshman

	People's first-class scholarship
2003	The third prize at China Olympic Math Competition for Middle School Students
2002	The second prize at the 22th National Physics Competition

Professional Skills

Languages
English, Chinese
Programming languages and tools
Python, C/C++, Java, MATLAB/Octave, Linux, GTK+, OpenCV, Lua, Cocos2d-X
Development Platforms
Microsoft Visual Studio, Eclipse, GNU/Linux

Publications

1. Xiangmin Fan, **Wencan Luo**, Jingtao Wang, Mastery Learning of Second Language through Asynchronous Modeling of Native Speakers in a Collaborative Mobile Game, Proceedings of ACM Conference on Human Factors in Computing Systems (CHI 2017), Denver, CO. (to appear)

2. Xiangmin Fan, **Wencan Luo**, Muhsin Menekse, Diane Litman, Jingtao Wang, *Scaling Reflection Prompts in Large Classrooms via Mobile Interfaces and Natural Language Processing*, In *Proceedings of 22nd ACM Conference on Intelligent User Interfaces (IUI 2017)*, Limassol, Cyprus.

3. **Wencan Luo**, Fei Liu, and Diane Litman, 2016, *An Improved Phrase-based Approach to Annotating and Summarizing Student Course Responses*, In *Proceedings of the 26th International Conference on Computational Linguistics (COLING 2016)*, Osaka, Japan.

4. **Wencan Luo** and Fan Yang, 2016, *An Empirical Study of Automatic Chinese Word Segmentation for Spoken Language Understanding and Named Entity Recognition*, In *Proceedings of the 15th Conference of the North American Chapter of the Association for Computational Linguistics – Human Language Technologies (NAACL-HLT 2016)*, San Diego, CA.

5. **Wencan Luo**, Fei Liu, Zitao Liu and Diane Litman, 2016, *Automatic Summarization of Student Course Feedback*, In *Proceedings of the 15th Conference of the North American Chapter of the Association for Computational Linguistics – Human Language Technologies (NAACL-HLT 2016)*, San Diego, CA.

6. **Wencan Luo** and Diane Litman, 2016, *Determining the Quality of a Student Reflective Response*, In *Proceedings 29th International FLAIRS Conference*, Key Largo, FL, May. (**Best Student Paper Award Nominee**)

7. **Wencan Luo** and Diane Litman. 2015. *Summarizing Student Responses to Reflection Prompts*, In

Proceedings Empirical Methods in Natural Language Processing (EMNLP), Lisbon, Portugal.

8. **Wencan Luo**, Xiangmin Fan, Muhsin Menekse, Jingtao Wang, and Diane Litman. 2015. *Enhancing Instructor-Student and Student-Student Interactions with Mobile Interfaces and Summarization*, In *Proceedings NAACL HLT Companion Volume*, pp. 16-20, Denver, Colorado, June. (demo)

9. Xiangmin Fan, **Wencan Luo**, Muhsin Menekse, Diane Litman, and Jingtao Wang. 2015. *CourseMIRROR: Enhancing Large Classroom Instructor-Student Interactions via Mobile Interfaces and Natural Language Processing*, In *Proceedings ACM CHI (Works-in-Progress)*, pp. 1473-1478, Seoul, Korea, April.

10. Ricardo Usbeck, Axel-Cyrille Ngonga Ngomo, **Wencan Luo**, Lars Wesemann. 2014. *Multilingual Disambiguation of Named Entities Using Linked Data*, In *Proceedings of the 13th International Semantic Web Conference*, Demos & Posters.

11. Zhongguo Yuan, **Wencan Luo**, Feng Tian, 2014, *Badminton training analysis system based on physiological computing*, *Computer Engineering and Applications 2014*, Issue 12, pp. 59-65

12. **Wencan Luo**, Joel Chan and Diane Litman. 2013. *Reducing Annotation Effort on Unbalanced Corpus based on Cost Matrix*, In *Proceedings of the NAACL HLT Student Research Workshop (NAACL-SRW 2013)*, Atlanta, GA, June.

13. Fei Lv, Feng Tian, Yingying Jiang, Xiang Cao, **Wencan Luo**, Guang Li, Xiaolong Zhang, Guozhong Dai, Hongan Wang. 2011. *ShadowStory: Creative and Collaborative Digital Storytelling Inspired by Cultural Heritage*, In *Proceedings of ACM Conference on Human factors in Computing Systems (CHI 2011)*, Pages: 1919-1928.

14. Feng Tian, Fei Lv, Jingtao Wang, Hongan Wang, **Wencan Luo**, Matthew Kam, Vidya Setlur, Guozhong Dai, John Canny. 2010. *Let's play chinese characters: mobile learning approaches via culturally inspired group games*, In *Proceedings of ACM Conference on Human factors in Computing Systems (CHI 2010)*, Pages: 1603-1612.

15. Feng Tian, **Wencan Luo**, Jing Dai, Yongmin Cheng, Hongan Wang, Guozhong Dai. 2010. *Leveraging Psychophysical Data in Monitoring and Analyzing the States of Badminton Players*, *CHI 2010 Workshop on Brain, Body and Bytes: Psychophysiological User Interaction*

16. Fei Lv, **Wencan Luo**, Feng Tian, Hongan Wang, Guozhong Dai. 2010. *UEMM: A New Method for the Evaluation of Reality Based Interaction*, In *Proceeding of the 6th Joint Conference on Harmonious Human Machine Environment (HHME 2010)*.