

yanchuan sim

machine learning and language technologies specialist

contact

1 fusionopolis way
singapore 138632

+65 9179 5907
ysim@cs.cmu.edu
http://yanchuan.sg

natural languages

english, mandarin,
conversational spanish

programming languages

python, c/c++, scala,
java, php, html/css,
javascript, ruby on
rails, sql, matlab

tools & frameworks

tensorflow, spark,
django, numpy, scipy,
scikit-learn, boost,
openmp, hadoop,
elastic mapreduce

awards

National Science
Scholarship (2006),
National Olympiad in
Informatics (2004,
Gold)

interests

snowboard, traveling,
hiking, tennis, aviation,
food!

summary

Current scientist in national research lab I²R with over 6 years of experience at the forefront of cutting edge NLP and ML research; co-authored papers with over 150 citations in top peer-reviewed conferences. Worked with multiple startups to capitalize on their data, garnering insights to improve product and business metrics. Super nerd who's been coding for 18+ years with an innate curiosity for exploring and experimenting with new technologies.

experience

- 2016–Now **Institute for Infocomm Research (I²R)** Fusionopolis, Singapore
Scientist, Natural Language Processing Department
I lead a team of NLP engineers to design and develop research prototypes for our industry partners. Our team has deep capabilities in a range of diverse information extraction tasks, such as **relation extraction**, **entity linking**, **sentiment analysis**, and **topic detection**.
- May-Aug '15 **AirPR Inc** San Francisco, CA
Software Engineering Intern / Technical Advisor
I collaborated with the engineering team to integrate modern NLP methods to AirPR's analytics product. Some key projects are: **automatic key phrase extraction**, topic analysis, and **author expertise ranking** algorithms, and social media **sentiment analysis**. We used Spark+AWS' EMR to efficiently process millions of web articles everyday.
- May-Aug '13 **Google Inc** Mountain View, CA
PhD Research Intern, Google Knowledge
As part of Google Knowledge's R&D team, we were tasked with pushing the boundaries of the state of the art NLP pipeline (chunking, tagging, coreference, etc). I proposed and implemented a novel model for performing joint inference on **entity recognition** and **coreference resolution**, developing efficient inference algorithms to find good solutions in the high dimensional combinatorics space using dual decomposition.

education

- 2011–2016 **PhD Language Technologies** Carnegie Mellon University
Advisor: Noah Smith, Noah ARK's Lab.
My research interests are **natural language processing** and building **statistical models** for applications in text mining and knowledge discovery. My thesis is titled "Text as a Strategic Choice", which I proposed a novel computational framework for studying how text is used strategically to influence behaviors. My coursework include NLP, probabilistic graphical models, advanced optimization, statistical machine learning & deep learning, analysis of social media.
- 2007–2010 **BSc Computer Science with Linguistics minor, Highest Honors** UIUC
GPA: 3.87/4.0. Thesis: *Inducing lexical clusters from unannotated corpora using CCG parsing* (advised by Prof. Julia Hockenmeier).