Lifeng Jia

|  |  |
| --- | --- |
| **ADDRESS** | **PHONE**: (312) 508-9599 |
| 2550 Waterview Dr, Apt 249 | **EMAIL**: jialifeng@gmail.com |
| Northbrook, IL, 60062 | **CITIZENSHIP:** U.S. Permanent Resident |

## EDUCATION

|  |  |
| --- | --- |
| **University of Illinois at Chicago**, USA  PhD of Computer Science | Aug 2006 - Aug 2013 |
| **Jilin University**, China  Master of Computer Science | Aug 2003 - May 2006 |
| **Jilin University**, China  Bachelor of Computer Science | Aug 1999 - May 2003 |

## PROFESSIONAL SKILLS

|  |
| --- |
| * Solid understanding of deep learning on: 1) CNN architectures and RNN architectures, such as ResNet and LSTM; 2) Attention mechanisms in encoder-decoder architectures; 3) Practical training techniques, such as drop out and batch normalization; 4) experiences on using deep learning packages, such as Keras , |
| * Deep understanding of machine learning on: 1) Parametric Bayesian models, such as LDA and Naive Bayes; 2) Discriminative Classification models, such as Logistic Regression, Neutral Network and SVM; 3) Non-Negative Matrix Factorization models, such as collaborative-filtering; 4) Ensemble learning, such as bootstrap and boosting; |
|  |
| * 4) Deep learning applications on natural language processing, such as sentiment analysis; |
| * using deep learning and machine learning packages, such as Keras and Scikit-Learn. |
| * Proficient in Python programming and Scala programming over Spark. |
| * Rich experiences in Spark programming over big data. |
| * Rich experiences on team management and excellent communication skills. |
| * Great enthusiasm for innovation and strong learning and communication skills. |

## EXPERIENCES

|  |  |  |
| --- | --- | --- |
| * **Machine Learning R&D Engineering** | **Networked Insights Inc.** | Aug 2012 – Aug 2015 |
| * **Social Media User-Level Demographic Classification** | | Jan 2015 – Aug 2015 |

Proposing models to classify demographic information (i.e. gender, age and geo) of social media users based on their social media posts. Bootstrapping post-level demographic classification results into user-level information. Efficiently learning classifiers by employing Stochastic Gradient Descent training with hinge loss and elastic net regularization. Implementing Hive-based UDFs and UDAFs within Hadoop framework to efficiently classify demographic information.

|  |  |
| --- | --- |
| * **Social Media Audience Creation for Digital Targeting** | Aug 2014 - Dec 2014 |

Designed an audience creation system of building sufficiently large audiences for digital targeting. The system is characterized by: 1) learning collaborative-filtering models to recommend new topical interests to social media users based on the original interests of their posts; 2) calculating interest profiles of given small seeds of users by statistical measures, such as adjusted residuals, z-score of odd-ratios and chi-square scores; 3) finding users sharing similar interest profiles to given seeds. Implemented the system by Hadoop and Apache Mahout.

|  |  |
| --- | --- |
| * **Consumer vs. Non-Consumer Twitter User Classification** | Feb 2014 - Jul 2014 |

Designed a classification system that differentiates consumer Twitter users from non-consumer Twitter users based on their monthly behaviors. Non-consumer Twitter users normally refer to spammers, bots, promoters, official Twitter users and so on. Implemented the model by defining Hive UDF/UDAFs within Hadoop framework. The system classifies 1.5 billion tweets within 1 hour.

|  |  |
| --- | --- |
| * **Social Media Spam Classification** | Aug 2013 - Jan 2014 |

Designed classifiers to identify spam posts within various social media feeds, such as Twitter, blogs and forums. Implemented classifiers by learning Random Forests and Naive Bayes models from Weka and BM25 model.

|  |  |
| --- | --- |
| * **Social Media Post-Level Demographic Classification** | Aug 2012 - Jul 2013 |

Designed models that classify the demographic information of social media users based on a single social media post. Implemented models by employing Naive Bayes model classifier Java Weka package, SGDClassifier from python Scikit-learn package and so on. Studied the precision-recall curves of learned models to guarantee high precisions in expense of recalls. For example, gender classifier achieves at least 80% in precision but at a recall of 72% while age classifier achieves the same precision at a recall of 20%.

|  |  |  |
| --- | --- | --- |
| * **Research Assistant** | **University of Illinois at Chicago** | Aug 2008 - May 2012 |
| * **Twitter Real-time Ad-hoc Retrieval System** | | Aug 2011 - May 2012 |

Proposed and developed a Twitter search system which is characterized by a divide-and-conquer framework of learning to rank tweets. Implemented the system by writing more than 40K lines of programs in Perl. Evaluated the system by TREC Tweets2011 collection consisting of about 16 million tweets. The system outperforms the best known results reported by TREC 2011.

|  |  |
| --- | --- |
| * **Text Mining Pipeline for Systematic Reviews** | Jan 2011 - May 2012 |

Participated in the development of a meta-search engine for medical literature which incorporates five individual search engines: PubMed, Embase, Cochrane, CINAHL and PsycInfo.

|  |  |
| --- | --- |
| * **Faceted Blog Distillation System** | Aug 2010 - Aug 2011 |

Proposed novel hierarchical Bayesian models to retrieve blogs that are not only relevant to queries but also exhibit facets, such as opinionated vs. factual, personal vs. official and in-depth vs. shallow. Trained the models by collapsed Gibbs sampling algorithm. Implemented the systems by writing more than 18K lines of programs in Perl. Evaluated them by TREC Blogs08 collection consisting of about 28.5 million blog posts and 1.3 million blogs. The systems outperforms the best known results reported by TREC 2010.

|  |  |
| --- | --- |
| * **Opinion Retrieval System** | Aug 2008 - Aug 2010 |

Proposed and developed an opinion retrieval system which retrieves the opinionated relevant blog posts. Implemented the system by writing more than 5K lines of Perl programs and 7K lines of Java programs. Evaluated it by TREC Blogs06 collection consisting of about 3.2 million blog posts and the system outperforms the state-of-the-art techniques.

|  |  |
| --- | --- |
| * **Negation Scope Detection System** | Aug 2008 - May 2009 |

Introduced the definition of the scope of negation. Proposed a system to calculate the exact scopes of negations within the context of sentences. Studied the impact of the scopes of negation on sentiment analysis. Implemented the system by writing more than 15K lines of programs in Java and evaluated it by TREC Blogs06 collection. The system outperforms the state-of-the-art techniques in terms of the accuracy of determining the scopes of negations and demonstrates the enhancement in the performance of sentiment analysis.

|  |  |
| --- | --- |
| * **Drug Name Confusion Detection** | Aug 2006 - Dec 2008 |

Developed a drug name search engine to retrieve drug names which have high confusability with a given query drug. Implemented the search engine by writing more than 10K lines of programs in Java.

## SELECTED PUBLICATIONS

|  |
| --- |
| * Neil R. Smalheiser, Can Lin, Lifeng Jia, Aaron M. Cohen, Clement Yu, John M. Davis, Clive E. Adams, Marian S. McDonagh and Weiyi Meng. "Design and Implementation of Metta, a Metasearch Engine for Biomedical Literature Retrieval Intended for Systematic Reviews". Health Information Science and Systems, Volume 2, Issue 1, Jan. 2014. |
| * Lifeng Jia, Clement Yu and Weiyi Meng. "The Impacts of Structural Difference and Temporality of Tweets on Retrieval Effectiveness". ACM Transactions on Information Systems (ACM TOIS), 2013. |
| * Lifeng Jia, Clement Yu and Weiyi Meng. "Faceted Models of Blog Feed". In Proceedings of the 22nd ACM International Conference on Information and Knowledge Management (CIKM), 2013. |
| * Lifeng Jia, Clement Yu, Weiyi Meng and Lei Zhang. "Facet-Driven Blog Feed Retrieval". In Proceedings of the 14th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing), 2013. |
| * Lifeng Jia and Clement Yu. "UIC at TREC 2010 Faceted Blog Distillation". In Proceedings of the 19th Text REtrieval Conference (TREC), 2010. |
| * Lifeng Jia, Clement Yu and Weiyi Meng. "The Effect of Negation on Sentiment Analysis and Retrieval Effectiveness". In Proceedings of the 18th ACM International Conference on Information and Knowledge Management (CIKM), 2009. |
| * Wei Zhang, Lifeng Jia, Clement Yu and Weiyi Meng. "Improve the Effectiveness of the Opinion Retrieval and Opinion Polarity classification". In Proceedings of the 17th ACM International Conference on Information and Knowledge Management (CIKM), 2008. |
| * Lifeng Jia, Clement Yu and Wei Zhang. "UIC at TREC 2008 Blog Track". In Proceedings of the 17th Text REtrieval Conference (TREC), 2008 |

## PROFESSION ACTIVITIES

|  |
| --- |
| * Invited Reviews for Journals: ACM TOIS, IEEE TKDE, JASIST, ACM TKDD, ACM TIST etc. |
| * Reviewers for Conferences: ICDM'14, ICWSM'14, EACL'14, WWW'14, PAKDD'13, EMNLP'13 etc. |

## HONORS AND AWARDS

|  |
| --- |
| * Presenter's Travel Award, University of Illinois at Chicago in 2008, 2010 and 2013. |
| * Graduate Student Council Award, University of Illinois at Chicago in 2008,2010 and 2013. |
| * Excellent Student Awards, Jilin University in 2000, 2001, 2002 and 2003. |