Lifeng Jia

ADDRESS
2550 Waterview Drive, Apt 249
Northbrook, IL, 60062

PHONE: (312) 508-9599
EMAIL: jialifeng@gmail.com
CITIZENSHIP: U.S. Permanent Resident

OBJECTIVE

Seeking a position related with natural language processing and optimization.

PROFESSIONAL SKILLS

- Solid understanding of deep learning on: 1) CNN and RNN architectures, such as ResNet and LSTM; 2) attention mechanisms in encoder-decoder architectures; 3) regularization techniques, such as drop out and batch normalization; 4) deep learning applications on NLP, such as sentiment analysis.
- Deep understanding of machine learning on: 1) supervised classification methods, such as Naive Bayes and SVM; 2) unsupervised clustering methods, such as K-means and DB-scan. 3) dimension reduction techniques, such as PCA; 4) ensemble learning methods, such as Bootstrap and Boosting;
- Proficient on Python, Scala and Spark programming and extensive practice in on machine learning packages, such as keras and scikit-learn.
- Great enthusiasm for innovation, strong motivation for learning and excellent management and communication skills.

EDUCATION

University of Illinois at Chicago, USA — 2006 - 2013
PhD of Computer Science

Jilin University, China — 2003 - 2006
Master of Computer Science

Jilin University, China — 1999 - 2003
Bachelor of Computer Science

EXPERIENCES

- **Director of Optimization** — **Conversant LLC.** — Aug 2015 – Present

➢ **Real-Time Bid Optimization Suite**

Built a suite of online models that optimize digital advertisement campaigns' performance in real time. These online models adjust campaigns' audiences (who to show ads), allocate budgets (how much money to spend on ads) and calculate bidding prices (how to spend budgets) to optimize campaigns' KPIs, such as conversions, viewability and clicks and so on.

➢ **Real-Time Auto-Pacing**

Established online models that calculate optimal bid prices in real-time for a digital advertisement campaign to efficiently spend its budgets. It is characterized by 1) a LSTM model to predict campaign's future online ads inventories; 2) a quadric programming model that allocates its total budgets over hours to minimize costs and 3) a pricing model to calculate bid prices in real-time to spend allocated budgets efficiently and effectively.

➢ **Bid Strategy Optimization**

Designed a model to calculate optimal long-term (daily or weekly) bidding strategies for digital advertisement campaigns. Specifically, the model first studies various campaigns' distributions (e.g. win rate vs. costs, probability of conversions) from their historical data and then computes their optimal bidding strategies to satisfy campaigns' performance goals while minimize their costs.

- **Machine Learning R&D Engineering** **Networked Insights Inc.** Aug 2012 – Aug 2015

➢ **Social Media User-Level Demographic Classification**

Proposed models to classify demographic information (i.e. gender, age and geo) of social media users based on their social media contents and meta attributes, such as tweets. Bootstrapping post-level demographic classification results into user-level information.

➢ **Social Media Audience Creation for Digital Targeting**

Designed an audience system of automatically creating sufficiently large audiences for digital advertisement campaigns. 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.

➢ **Consumer vs. Non-Consumer Twitter User Classification**

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. The system classifies 1.5 billion tweets within 1 hour.

➢ **Social Media Spam Classification**

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**

Designed models that classify the demographic information of social media users based on a single social media post. Implemented models by employing 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 with 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 - Aug 2012

➢ **Twitter Real-time Ad-hoc Retrieval System**

Proposed and developed a Twitter search system which is characterized by a divide-and-conquer framework of learning to rank tweets. Evaluated the system by TREC Tweets2011 collection of 16 million tweets. The system outperforms the best-known results reported by TREC 2011.

➢ **Faceted Blog Distillation System**

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. Evaluated them by TREC Blogs08 collection of 28.5 million blog posts and 1.3 million blogs. The systems outperform the best-known results reported by TREC 2010.

➢ **Negation Impacts on Sentiment Analysis**

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. 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 significant improvements on sentiment analysis.

- **Opinion Retrieval System**

Proposed and developed an opinion retrieval system which retrieves the opinionated relevant blog posts. Evaluated it by TREC Blogs06 collection of 3.2 million blog posts and the system outperforms the state-of-the-art techniques.

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.

CERTFICIATIONS

Deep Learning Specialization — Coursera

- Neural Networks and Deep Learning
- Structuring Machine Learning Projects
- Improving Deep Neural Networks: Hyperparameter tuning, Regularization and Optimization
- Convolutional Neural Networks
- Sequence Models

Functional Programming in Scala Specialization — Coursera

- Functional Programming Principles in Scala
- Functional Program Design in Scala
- Parallel programming
- Big Data Analysis with Scala and Spark
- Functional Programming in Scala Capstone

Basic Modeling for Discrete Optimization — Coursera