

# Fei Sun

## Curriculum Vitae

Alimama  
Alibaba Group  
Building No.9 Wangjing East Garden 4th Area  
Chaoyang District, Beijing, China, 100102  
☎ 13488810349  
✉ ofey.sunfei@gmail.com  
📄 ofey.me  
👤 FeiSun

### WORK EXPERIENCE

2017– Senior Algorithm Engineer, Alibaba, Beijing.

### RESEARCH INTEREST

NLP Distributed Representation, Text Representation  
Machine Learning Learning Algorithms, Deep Learning

### EDUCATION

- 2012–2017 **Institute of Computing Technology, Chinese Academy of Sciences.**  
*Ph.D. in Computer Science*  
– Research Advisor: Prof. Jun Xu, Prof. Jiafeng Guo, & Prof. Xueqi Cheng
- 2009–2012 **Beijing Institute of Technology, School of Computer Science.**  
*M.S. in Computer Science*  
– Research Advisor: Prof. Lejian Liao
- 2005–2009 **Beijing Institute of Technology, School of Software.**  
*BEng in Software Engineering*  
– Rank: No.9 out of 213 undergraduates in school of software

### RESEARCH EXPERIENCE

#### Word Representation

- 10/2015–02/2016 **Interpretable Word Representations [1]**, *Institute of Computing Technology, IJCAI 2016 Oral.*  
A critical issue for dense word embeddings is that they are lacking of interpretability. To tackle this issue, we introduced the sparse constraint to Word2Vec and used regularized dual averaging algorithm to optimize it in online training.
- 06/2015–09/2015 **Distributed Word Representations and Phrase Representations [2]**, *Institute of Computing Technology, AAAI 2016 Oral.*  
Most existing word representation models based on distributional hypothesis cannot handle rare words well. In this work, we proposed two novel models to learn better word representations using both external contexts and internal morphemes. In addition, we also applied these two models to phrase representation.
- 08/2014–02/2015 **Distributed Word Representations [4]**, *Institute of Computing Technology, ACL 2015 Oral.*  
Existing distributed word representation methods modeled either syntagmatic or paradigmatic relations between words. In this work, we learned word representations using both syntagmatic and paradigmatic relations via a joint objective function.

#### Information Extraction

06–12/2010 **Web Content Extraction [6, 5]**, *Beijing Institute of Technology, SIGIR 2011 Oral & KAIS*.

We Proposed a DOM based content extraction approach via text density. The proposed approach improves the quality of structural content extraction of web pages and retains the original structural information in the web page cleaning process.

---

## SKILLS

Programming C/C++, Python, Matlab  
Operating System Linux/Unix, OS X, Windows  
Language Chinese (native), English (fluent)

---

## AWARDS & HONORS

02/2016 **AAAI Student Scholarship**.  
2015 **Inspur PhD Scholarship**.  
11/2012 **IBM China Excellent Student Scholarship**.  
12/2011 **Xu Teli Scholarship**.  
Greatest Honor of Beijing Institute of Technology, top 21/30,000 students.  
12/2011 **Pacesetter of Outstanding Graduate Students**.  
Awarded to the top 1% of each department.  
07/2009 **Outstanding Graduates of Beijing Institute of Technology**.  
Awarded to the top 10% of each department.

---

## HOBBIES

Reading History, Mathematics, Physics, Science Fiction, and Chinese Poetry  
Travel Interesting Places

---

## PROFESSIONAL ACTIVITIES

Service **PC Member**: EMNLP 2016, EACL 2017, EMNLP 2017  
Memberships ACL Student Membership, AAAI Student Membership  
Volunteer WSDM 2015, IJCAI 2016

---

## SELECTED PUBLICATIONS

**Fei Sun**, Jiafeng Guo, Yanyan Lan, Jun Xu, and Xueqi Cheng. Sparse Word Embeddings Using  $\ell_1$  Regularized Online Learning. In *Proceedings of IJCAI*, pages 2915–2921, 2016.

**Fei Sun**, Jiafeng Guo, Yanyan Lan, Jun Xu, and Xueqi Cheng. *Inside Out*: Two Jointly Predictive Models for Word Representations and Phrase Representations. In *Proceedings of AAAI*, pages 2821–2827, 2016.

Yanran Li, Wenjie Li, **Fei Sun**, and Sujian Li. Component-enhanced chinese character embeddings. In *Proceedings of EMNLP*, pages 829–834, 2015.

**Fei Sun**, Jiafeng Guo, Yanyan Lan, Jun Xu, and Xueqi Cheng. Learning word representations by jointly modeling syntagmatic and paradigmatic relations. In *Proceedings of ACL*, pages 136–145, 2015.

Dandan Song, **Fei Sun**, and Lejian Liao. A hybrid approach for content ex-

traction with text density and visual importance of dom nodes. *Knowledge and Information Systems*, pages 75–96, 2015.

**Fei Sun**, Dandan Song, and Lejian Liao. Dom based content extraction via text density. In *Proceedings of SIGIR*, pages 245–254, 2011.