Juan XU

4 years experience on NLP&ML

CONTACT INFORMATION

Tel: +86 13916430285
Email: feiathna@126.com

WORKING EXPERIENCE

MeiTuan-DianPing NLP Algorithm Engineer **Jul. 2014 – Present**

◆ **Query understanding**

Built a query understanding platform for better search result.

➢ **Semantic tagging:**

(a) Designed and implemented three methods on parsing vertical search queries that enables automatic tagging of the queries, the system consists of rule-based components, CRF and BiLSTM+CRF;
(b) Achieved 94% recall and 95% precision on DianPing's daily search queries.
(c) Technology Innovation Award Winner.

➢ **Query Rewrite** **Dec. 2016 - Present**

• **Query Correction**

(a) According to rules, such as pinyin, fuzzy pinyin, edit distance, English prefix matching, generate correction candidates from session data and offline Knowledge base.
(b) Trained a back off tri-gram language model on 12 million queries from DianPing search log;
(c) Ranking correction candidate using LR and the language model to choose the best accurate word.
(d) 1.05% queries were corrected by this system.

• **Query expansion**

(a) Built semantic dictionaries off-line, including synonym, hypernym and hyponym of search keyword.
(b) Developed semantic alignment techniques based on multi-granularity word-segments to ensure consistency of semantic intent between rewriting and original words.

• **Query term weight.**

(a) Segment search queries based on Chinese word segmentation method.
(b) Based on rules and CRF methods to weight the importance of the segmentation words, eliminate less important word.

➢ **Query category intent analysis** **Oct. 2015 - Present**

Predicated query related category for better ranking result.
(a) Predicated the second-level poi category of the query using click model and Bayesian text model。
(b) Designed and implemented CNN, RNN method to predict the first class of query.

◆ **Text Mining** **Jul. 2017 – Mar.2018**

Be responsible for text mining of UGC for content search.

➢ **Keyword extraction**

Using TF-IDF, Entity extraction, LDA and TextRank method to extract the key words of the article.

➢ **Opinion Mining:**

(a) Designed and implemented a LSTM+CRF based system for opinion mining;
(b) Nearly hundreds of millions of comments on shops in dianping.com have been processed by this system, and millions of opinion haven been extracted.

- **UGC Content Summary and article generation**

- **Content Summary**
  (a) Extracted the multiple topic dimension sentence based on rules and LSTM+CRF method. And we introduce the score of emotion into sentence filtering.
  (b) Ranked the candidate sentence based on TextRank method.

- **Title extraction**
  (a) Designed and implemented CNN, RNN method to extract title sentence of UGC.

- **Title Generation**
  (a) Extracted keywords of UGC sentence using entity extraction and other related techniques.
  (b) Designed and implemented an NMT method to generate sentence with the extracted keywords.

EDUCATION

Fudan University (Fudan) Shanghai, China
Supervisor Xuanjing Hunag QiZhang
Research area NLP
Master of Natural Science in Computer Applied Technology *Sep.2011-Jun.2014*

Anhui University of Technology (AHUT) Maanshan, China
Bachelor of Engineering in Software Engineering *Sep.2006-Jun.2010*

PUBLICATION & PATENT

- **Xu Juan**, Zhang Qi, Huang Xuanjing. **Understanding the Semantic Intent of Domain – Specific Natural Language Query**. International Joint Conference on Natural Language Processing, pages 552C560, Nagoya, Japan, 14-18-October 2013
- **Xu Juan**, Zhang Qi, Huang Xuanjing**. Personalized Hashtag Suggestion for Microblogs**, 4th National Conference, SMP 2015, Guangzhou, China, November 16-17, 2015, Proceedings， pp 38-50
- **A general method and system for identifying the semantic intent of search keyword**. 2017.12. Second class patent of meituan-dianping company.
- **A method and system of real-time generation of articles.** 2018.01. First class patent of meituan-dianping company.

SKILLS

Language: Chinese (Native), English (Fluent)
NLP: (CWS, PosTagging, Language Model, NER)
Deep Learning: (CNN, RNN, Attention, Embedding, Tensorflow, PyTorch)
Java: (Spring, Maven, Design Pattern)
Python: (numpy, scipy, pandas ,sklearn, beautiful Soup)