‍‍何云超

**研究方向：**文本情感分析，自然语言处理，深度学习，机器学习

**Email:** yunchaohe@gmail.com **Phone:** 15288217670

基本信息

|  |  |  |
| --- | --- | --- |
| 姓名 |  | 何云超 （男） |
| 学历 |  | 硕士研究生（云南大学、台湾元智大学双学位） |
| 研究方向 |  | 文本情感分析，自然语言处理，深度学习，机器学习 |
| 导师 |  | 禹良治教授（台湾）、刘惟一教授（大陆） |
| 学习经历 |  | 台湾元智大学、云南大学、天津工业大学 |
| 毕业时间 |  | 2016年5月 |
| 联系电话 |  | (+86) 15288217670 |
| Email |  | yunchaohe@gmail.com |
| GitHub |  | <https://github.com/candlewill> |
| 技术博客 |  | Python (<http://paisen.blogspot.com/>)  情感分析 (<http://sentiment-mining.blogspot.com/>) |

研究兴趣

* 自然语言处理
  + 单词级别和文本级别情感分析
  + 连续型情感强度预测
  + 文本情感倾向分类
  + 深度神经网络在自然语言处理中的应用 (E.g., word2vec, GloVe, sentence vectors)
* 机器学习
  + 传统机器学习算法
  + 深度学习(CNN, DNN, RNN, GRU, LSTM等)

教育背景

* **台湾元智大学** （第二学位，硕士研究生） 2015年3月 – 2016年2月

台湾元智大学，资讯学院资讯工程学系，自然语言处理与文字探勘实验室（导师：禹良治教授）

* **云南大学** （第一学位，硕士研究生） 2013年9月 – 2016年5月

云南大学，信息学院，计算机应用技术专业，数据与知识工程研究室（导师：刘惟一教授）

* **天津工业大学**（理学学士，本科） 2009年9月 – 2013年6月

天津工业大学，理学院，信息与计算科学专业

研究经历

* 情绪维度辨识与个人化文件推荐于社群网路家人关怀之研究 2015年5月 – 2016年2月

(台湾行政院国家科学委员会专题研究计划)

* + 利用回归模型拟合词向量和单词情感强度之间的关系(线性回归、SVR)
  + 构建词向量和句向量之间的非线性关系(CNN)
  + 使用深度学习方法分析文本情感强度(加权方法、CNN for regression)
* 协助导师撰写2016-2019年度项目计划书
  + 主要负责传统情感词典、情感语料库收集整理，以及词向量优化的部分
* 社群媒体分析于消费者健康资讯之应用 2015年2月 – 2015年6月

(台湾行政院国家科学委员会专题研究计划)

* + 利用聚类方法对Twitter短文本进行预处理以构建长文本
  + 分析非正式短文本的情感类别
* 基于概率图模型的海量评分数据分析与用户行为建模 2014年10月 – 2015年2月

(国家自然科学基金面上项目)

* + 探索独立因果路径条件下的贝叶斯网络快速推理算法

所获奖励

* 2016年度，第10届国际SemEval-2016 Task 4 Subtask C情感分析竞赛在micro-MAE度量上取得**第二名**

(新闻报道：[YZU](http://www.yzu.edu.tw/index.php/component/option,com_alphacontent/section,2/cat,1/task,view/id,11070/Itemid,236/lang,tw/)，[Sina](http://news.sina.com.tw/article/20160308/16309006.html)，[自由时报](http://news.ltn.com.tw/news/life/breakingnews/1624996))

* 2015年度，云南省研究生省政府**奖学金** (Top 5)
* 2015年度，美国数学建模竞赛成功参与奖 (ICM-2015)
* 2014年度，第十一届全国研究生数学建模竞赛**一等奖**
* 2013年度，天津市人民政府**奖学金** (Top 5)
* 2013年度，天津工业大学校长**一等**奖学金
* 2012年度，天津工业大学校长**二等**奖学金

论文发表

* Chin-Sheng Yang, Yung-Chih Chen, Yunchao He. (2015, July). MINING OPINION LEADERS IN SOCIAL MEDIA. *The Third Taiwan Summer Workshop on Information Management* (TSWIM 2016), Taipei, Taiwan.
* Yunchao He, Chin-Sheng Yang, Liang-Chih Yu, K. Robert Lai, Weiyi Liu. (2015, December). Sentiment Classification of Short Texts based on Semantic Clustering. *2015 International Conference on Orange Technologies* (ICOT 2015), Hong Kong, China. (Oral Presentation: [PPT](http://www.slideshare.net/YunchaoHe/yunchao-he-icot2015-59452942))
* Liang-Chih Yu, Lung-Hao Lee, Shuai Hao, Jun Hu and K. Robert Lai, Yunchao He. (2016, June). Building Chinese Affective Resources in Valence-Arousal Dimensions. *The 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (NAACL HLT 2016), San Diego, California. (Accepted) (负责实验部分)
* Yunchao He, Liang-Chih Yu, K. Robert Lai, Weiyi Liu. (2016). Word Embeddings Refinement Using Sentiment Lexicons. In *Proceedings of the 54rd Annual Meeting of the Association for Computational Linguistics* (ACL 2016), Berlin, Germany. (Under Review)
* Yunchao He, Liang-Chih Yu, Chin-Sheng Yang, K. Robert Lai, Weiyi Liu. (2016). YZU-NLP Team at SemEval-2016 Task 4: Ordinal Sentiment Classification Using a Recurrent Convolutional Network. In *Proceedings of the 10th International Workshop on Semantic Evaluation* (SemEval 2016 at NAACL), San Diego, California. (Accepted)
* Yunchao He, Liang-Chih Yu, Jin Wang, K. Robert Lai, Weiyi Liu. (2016). Sentence Level Sentiment Intensity Detecting using Convolutional Neural Networks. *The 26th International Conference on Computational Linguistics* (COLING 2016), Osaka, Japan. (In Progress)
* [硕士毕业论文] 何云超. (2016). 聚类算法和卷积神经网络在文本情感分析中的应用研究. 云南大学.
* [本科毕业论文] 何云超. (2013). 非织造布基于分形维数的物理特性的研究. 天津工业大学.

项目经验

* 主要实验 (Python) 2015年2月 – 2016年3月
  + Ordinal\_Classification 2016年1月 ([Source](https://github.com/candlewill/Ordinal_classification))

整合LSTM模型和CNN模型，构建了LSTM-CNN模型，将Twitter文本按照其情感强度划分为五点度量。

* + Vecamend 2015年11月 – 2016年2月 ([Source](https://github.com/candlewill/Vecamend-master2))

运用极性情感词典对现有词向量(包括word2vec和GloVe)进行优化，使其更加适用于情感分析。

* + Chinese\_word\_vectors 2016年1月 ([Source](https://github.com/candlewill/Chinsese_word_vectors))

使用简体和繁体维中文基百科语料库，基于word2vec和GloVe模型，训练中文词向量。

* + Concept\_word\_embeddings 2015年11月 ([Source](https://github.com/candlewill/Concept_word_embeddings))

利用WordNet同义词典，将句中词汇用其同义词集替换，训练doc2vec句向量，用回归模型预测其情感值(valence-arousal)。

* + Thesis\_experiment 2016年1月 – 2016年3月 ([Source](https://github.com/candlewill/Thesis_experiment))

利用CNN模型，使用预先训练好的词向量，预测短文本的情感强度

* + short\_texts\_sentiment\_analysis 2015年6月 ([Source](https://github.com/candlewill/short_texts_sentiment_analysis))

基于情感词典的方法，用情感词汇的多种组合方式(平均值、TF加权平均值、TF-IDF加权平均值、几何平均值、加权平均值)预测短文本的情感大小。

* 参加Kaggle-2015比赛，使用bag-of-words特征对IMDB电影评论数据的情感分类，结果排156名(共578队参赛)
* T9搜索 2014年7月 ([APK](http://shouji.baidu.com/soft/item?docid=6658785&from=as&f=search_app_T9%E6%90%9C%E7%B4%A2%40list_1_title%404%40header_all_input), [GitHub](https://github.com/candlewill/T9Search))

(主要开发者)

T9搜索是一个Android手机上的应用，主要功能是依据应用首字母快速定位用户想要打开的软件。具体而言，将26个英文字母映射到9个数字键（九宫格），用9个键代替26个，支持拼音首字母、全拼、多音字、部分拼音、混拼搜索，使用的核心算法为前缀树（Prefix Tree）。

相关能力

* 编程语言：Python (熟练)，C(熟练)，Java (熟练)，Matlab (熟练)，C#(了解)
* 英语：CET-6