‍‍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（学术型） |
| 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 |  |  |
| 应聘岗位 |  | 算法工程师 |

研究兴趣

* 自然语言处理
* 单词级别和文本级别情感分析
* 深度神经网络在自然语言处理中的应用 (e.g., word2vec, GloVe, sentence vectors)
* 连续型情感强度预测
* 文本情感倾向分类
* 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年度项目計劃書

主要负责传统情感词典、情感语料库收集，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 (2015.02 – 今)

(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 (2014.10 – 2015.02)

(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，SemEval-2016 Task 4 Subtask C情感分析竞赛micro-MAE度量第二名 (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
* 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
* 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). MINING OPINION LEADERS IN SOCIAL MEDIA. *Third Taiwan Summer Workshop on Information Management.*
* Yunchao He, Liang-Chih Yu, Jin Wang, K. Robert Lai, Weiyi Liu. (2015). Sentence Level Sentiment Intensity Detecting using Convolutional Neural Networks. *Thirtieth AAAI Conference on Artificial Intelligence*. (Under Review)
* Yunchao He, Chin-Sheng Yang, Liang-Chih Yu, K. Robert Lai2 and Weiyi Liu. (2015). Sentiment Classification of Short Texts based on Semantic Clustering. *2015 International Conference on Orange Technologies* (ICOT2015). (Under Review)
* 毕业论文

本科毕业论文：非织造布基于分形维数的物理特性的研究Physical Characteristics Studying of Non-woven Fabric based on the Fractal Dimension

继续加

项目经验

* T9搜索 2014年7月

(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）。

Demo: <http://goo.gl/Chfuil> GitHub: <https://github.com/candlewill/T9Search>

这是读研期间主要的代码，里面有：

\* 使用gensim训练词向量、句向量，分别对应word2vec和doc2vec

\* 计算文本情感得分的几种方法：平均值、TF加权平均值、TFIDF加权平均值、几何平均值、加权平均值

\* 以及一些CNN的结构，利用深度学习的方法，来计算文本的情感强度和兴奋程度(valence and arousal)

\* 最后包括一些正在使用的语料库，包括我们自己标记的CVAT，以及anew, extended anew等内容

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)