

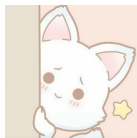
NLP最佳入门与提升路线

原创 夕小瑶 夕小瑶的卖萌屋 2月2日

来自专辑

卖萌屋@自然语言处理

>



一只小狐狸带你解锁NLP/ML/DL秘籍



作者：夕小瑶，小鹿鹿鹿，QvQ

前言

对突如其来的长假感到惶恐和不安？紧盯2019-nCoV的最新消息却依然感觉很空虚？腰酸萎靡脖子僵甚至怀疑自己有点发烧？**这是长时间没学习的症状。**



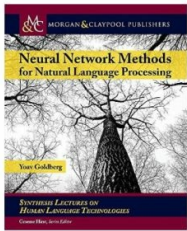
很久以前小夕推送的[深度学习入门资料推荐](#)和[机器学习与数学基础入门系列](#)收到到了很多小伙伴的好评和感谢，也因此一直有呼声希望小夕写一篇NLP方向的入门指导。于是，趁着这个长假，终于将这一篇拖了两年的稿子结了，希望能帮助大家哦。

入门圣经

一如既往的还是那句话，**不要养成囤书却不看书的习惯！！**入门阶段，精读一到两本经典书籍足矣。这里贴出来的书籍小夕在求学阶段都刷过，精读了这里的第1、2和4这三本（第二本由于太厚，到今天还没刷完），粗读了第3本，另外也粗刷了其他一些奇奇怪怪的书籍。

小夕这里吐血推荐前两本，第一本书适合快速入门，第二本书则是修炼内核的圣经书。两本书非常互补，入门阶段足够了。另外也给出了其他两本参考书，有兴趣的小伙伴根据自身精力阅读。

1. neural network methods for natural language processing



作者: Yoav Goldberg
出版社: Morgan & Claypool Publishers
出版年: 2017-4-17
页数: 310
定价: USD 74.95
装帧: Paperback
ISBN: 9781627052986

豆瓣评分



豆瓣书评传送门: <https://book.douban.com/subject/27032271/>

注: 订阅后台回复关键字【NLP入门书】可获取PDF下载链接

2. speech and language processing



作者: Daniel Jurafsky / James H. Martin
出版社: 人民邮电出版社
副标题: : 自然语言处理、计算语言学和语音识别导论
原作名: Speech and Language Processing
出版年: 2010-12-5
页数: 1024
定价: 138.00元
装帧: 平装
ISBN: 9787115238924

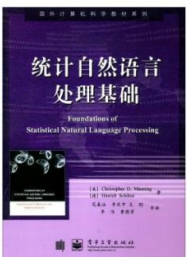
豆瓣评分



豆瓣书评传送门: <https://book.douban.com/subject/5373023/>

注: 订阅后台回复关键字【NLP入门书】可获取PDF下载链接

3. Foundations of Statistical Natural Language Processing



作者: Chris Manning / Hinrich Schütze
出版社: 电子工业出版社
原作名: Foundations of Statistical Natural Language Processing
译者: 苑春法 / 李伟 / 李庆中
出版年: 2005-1
页数: 418
定价: 55.00元
装帧: 平装(无盘)
丛书: 国外计算机科学教材系列
ISBN: 9787505399211

豆瓣评分



豆瓣书评传送门: <https://book.douban.com/subject/1224802/>

注: 订阅后台回复关键字【NLP入门书】可获取PDF下载链接

4. 统计自然语言处理



作者: 宗成庆
出版社: 清华大学出版社
出版年: 2013-8-1
页数: 570
定价: 85
装帧: 平装
丛书: 中文信息处理丛书
ISBN: 9787302319115

豆瓣评分



豆瓣书评传送门: <https://book.douban.com/subject/25746399/>

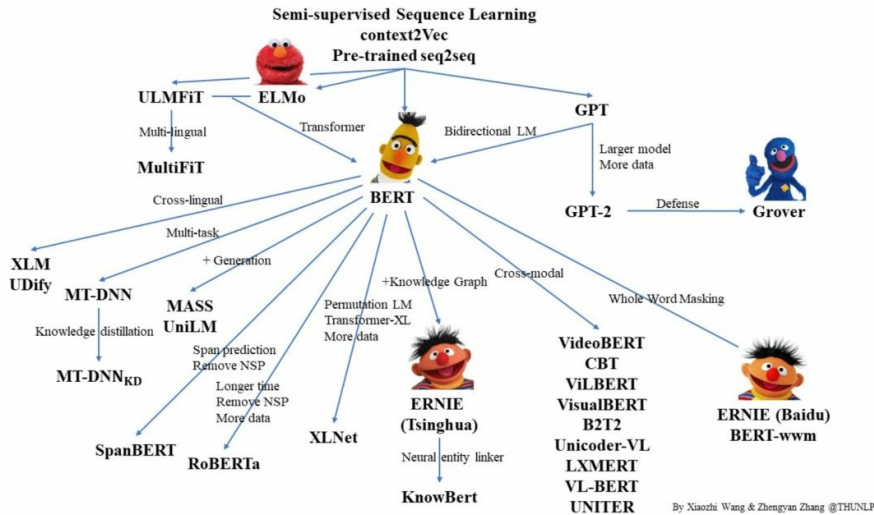
进阶打卡

对于已经入门NLP的小伙伴, 自然要paper刷起啦! 还在自己苦苦搜集最新paper?

想知道某个NLP方向的最新进展？贴心的小夕已经为大家准备好啦(￣▽￣)

注：由于每个方向的list都比较长，因此本文只列出每个方向的前五篇文章，**每个方向的完整list**请通过**对应的关键词**在**订阅号后台**领取哦。

pretrain models 【预训练模型】



1. HUBERT Untangles BERT to Improve Transfer across NLP Tasks. Anonymous authors. ICLR 2020 under review.
2. ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators. Anonymous authors. ICLR 2020 under review.
3. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, Peter J. Liu. Preprint.
4. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, Luke Zettlemoyer. Preprint.
5. MultiFiT: Efficient Multi-lingual Language Model Fine-tuning. Julian Eisenschlos, Sebastian Ruder, Piotr Czapla, Marcin Kardas, Sylvain Gugger, Jeremy Howard. EMNLP 2019.

注：完整版最新paper list在后台回复【预训练】领取（持续更新完善ing）

Machine Translation (MT) 【机器翻译】

1. Effective adversarial regularization for neural machine translation. Sano, Motoki, Jun Suzuki, and Shun Kiyono. ACL 2019.
2. Robust neural machine translation with doubly adversarial inputs. Cheng, Yong, Lu Jiang, and Wolfgang Macherey. ACL 2019.
3. Learning deep transformer models for machine translation. Wang, Qiang, et al. ACL 2019.
4. Self-Supervised Neural Machine Translation. Ruiter, Dana, Cristina Espana-Bonet, and Josef van Genabith. ACL 2019.
5. Bridging the gap between training and inference for neural machine translation. Zhang, Wen, et al. ACL 2019.

注：完整版最新paper list在后台回复【机器翻译】领取（持续更新完善ing）

Question Answer Systems 【问答系统】

1. NumNet: Machine Reading Comprehension with Numerical Reasoning. EMNLP 2019.
2. What's Missing: A Knowledge Gap Guided Approach for Multi-hop Question Answering. EMNLP 2019.
3. BiPaR: A Bilingual Parallel Dataset for Multilingual and Cross-lingual Reading Comprehension Novels. EMNLP 2019
4. Read + Verify: Machine Reading Comprehension with Unanswerable Questions. Minghao Hu, Furu Wei, Yuxing Peng, Zhen Huang, Nan Yang, and Ming Zhou. AAAI 2019.
5. Language Models as Knowledge Bases? EMNLP 2019

注：完整版最新paper list在后台回复【问答系统】领取（持续更新完善ing）

Dialogue Systems 【对话系统】

1. Sarik Ghazarian, Weischedel Ralph, Aram Galstyan, Nanyun Peng. Predictive Engagement: An Efficient Metric for Automatic Evaluation of Open-Domain Dialogue Systems AAAI 2020.
2. Exploiting Persona Information for Diverse Generation of Conversational Responses. IJCAI, 2019.
3. Global-to-local Memory Pointer Networks for Task-Oriented Dialogue. ICLR, 2019.
4. Towards Scalable Multi-domain Conversational Agents: The Schema-Guided Dialogue Dataset. 2019
5. Fan Wang Jinhua Peng Hua Wu Rongzhong Lian, Min Xie. Learning to select knowledge for response generation in dialog systems. arXiv preprint arXiv:1902.04911, 2019.

注：完整版最新paper list在后台回复【对话系统】领取（持续更新完善ing）

Data augmentation 【数据增强】

1. Xie Q , Dai Z , Hovy E , et al. Unsupervised Data Augmentation for Consistency Training[J]. 2019.
2. Ho D , Liang E , Stoica I , et al. Population Based Augmentation: Efficient Learning of Augmentation Policy Schedules[J]. 2019.
3. Cubuk, Ekin D, Zoph, Barret, Mane, Dandelion,等. AutoAugment: Learning Augmentation Policies from Data[J].
4. Kobayashi S . Contextual Augmentation: Data Augmentation by Words with Paradigmatic Relations[J]. 2018.
5. Wang X, Pham H, Dai Z, et al. Switchout: an efficient data augmentation algorithm for neural machine translation[J]. arXiv preprint arXiv:1808.07512, 2018.

注：完整版最新paper list在后台回复【数据增强】领取（持续更新完善ing）

Textual Adversarial Attack and Defense 【文本对抗样本与防御】

1. Analysis Methods in Neural Language Processing: A Survey. Yonatan Belinkov, James Glass. TACL 2019.
2. Towards a Robust Deep Neural Network in Text Domain A Survey. Wenqi Wang, Lina Wang, Benxiao Tang, Run Wang, Aoshuang Ye. 2019.
3. Generating Natural Language Adversarial Examples through Probability Weighted Word Saliency. ACL 2019.
4. Universal Adversarial Triggers for Attacking and Analyzing NLP. EMNLP-IJCNLP 2019.
5. LexicalAT: Lexical-Based Adversarial Reinforcement Training for Robust Sentiment Classification. EMNLP-IJCNLP 2019.

注：完整版最新paper list在后台回复【对抗样本】领取（持续更新完善ing）

Graph Neural Networks 【图神经网络】

1. A Comprehensive Survey on Graph Neural Networks. arxiv 2019.
2. DyRep: Learning Representations over Dynamic Graphs. ICLR 2019.
3. Hypergraph Neural Networks. AAAI 2019.
4. Graph Neural Networks: A Review of Methods and Applications. arxiv 2018.
5. Self-Attention Graph Pooling. ICML 2019.

注：完整版最新paper list在后台回复【图神经网络】领取（持续更新完善ing）

Style Transfer 【文本风格迁移】

1. A Dual Reinforcement Learning Framework for Unsupervised Text Style Transfer. Fuli Luo. IJCAI-2019.
2. Style Transfer in Text: Exploration and Evaluation, Zhenxin Fu from PKU, AAAI-2018.
3. Fighting Offensive Language on Social Media with Unsupervised Text Style Transfer. Cicero Nogueira dos Santos, ACL-2018.
4. Style Transfer from Non-Parallel Text by Cross-Alignment, NIPS-2017, Tianxiao Shen.

5. Unsupervised Text Style Transfer using Language Models as Discriminators, Zichao Yang, Arxiv.

注：完整版最新paper list在后台回复【风格迁移】领取（持续更新完善ing）

Knowledge Graph & Knowledge Representation 【知识图谱与知识表示】

1. RSN: Learning to Exploit Long-term Relational Dependencies in Knowledge Graphs. Lingbing Guo, Zequn Sun, Wei Hu. ICML 2019.
2. DihEdral: Relation Embedding with Dihedral Group in Knowledge Graph. Canran Xu, Ruijiang Li. ACL 2019.
3. CapsE: A Capsule Network-based Embedding Model for Knowledge Graph Completion and Search Personalization. Dai Quoc Nguyen, Thanh Vu, Tu Dinh Nguyen, Dat Quoc Nguyen, Dinh Q. Phung. NAACL-HIT 2019.
4. CaRe: Open Knowledge Graph Embeddings Swapnil Gupta, Sreyash Kenkre, Partha Talukdar. EMNLP-IJCNLP 2019.
5. Representation Learning: A Review and New Perspectives. Yoshua Bengio, Aaron Courville, and Pascal Vincent. TPAMI 2013.

注：完整版最新paper list在后台回复【知识图谱】领取（持续更新完善ing）

Information Extraction 【信息抽取】

3. FewRel: A Large-Scale Supervised Few-Shot Relation Classification Dataset with State-of-the-Art Evaluation. EMNLP 2018.
4. Hierarchical Relation Extraction with Coarse-to-Fine Grained Attention Xu Han, Pengfei Yu, Zhiyuan Liu, Maosong Sun, Peng Li. EMNLP 2018.
5. RESIDE: Improving Distantly-Supervised Neural Relation Extraction using Side Information Shikhar Vashishth, Rishabh Joshi, Sai Suman Prayaga, Chiranjib Bhattacharyya, Partha Talukdar. EMNLP 2018.
1. A Survey of Deep Learning Methods for Relation Extraction. Shantanu Kumar. 2017.
2. Relation Extraction : A Survey. Sachin Pawara,b, Girish K. Palshikara, Pushpak Bhattacharyya. 2017.

注：完整版最新paper list在后台回复【信息抽取】领取（持续更新完善ing）

当然啦，NLP的研究方向远不止这些，像词法分析、句法分析、语义分析等NLP基础问题，以及信息检索、文本摘要、篇章理解等重要应用方向暂时没有在这里贴出。后续本文会在订阅号主页下方「知识网络」标签页内持续更新，不断完善和跟进已有方向的paper list，增加新的大方向和小方向，敬请期待哦～

欢迎大家分享学习，让更多有需要的小伙伴看到～你的【在看】也是对小夕的鼓励♥

可能喜欢

- 深度学习资料挑的眼花啦？小夕帮你做选择！
- 论文写作葵花宝典
- 机器学习与数学基础入门手册（一）
- 机器学习与数学基础入门手册（二）
- 机器学习与数学基础入门手册（三）
- 计算机基础算法入门手册
- 万万没想到，我的炼丹炉坏了



夕小瑶的卖萌屋

关注&星标小夕，带你解锁AI秘籍



文章已于修改

声明：pdf仅供学习使用，一切版权归原创公众号所有；建议持续关注原创公众号获取最新文章，学习愉快！