

先人一步，敲开自然语言 处理这扇门

 犀牛学院  网易云课堂 |  微专业

目录

CONTENTS

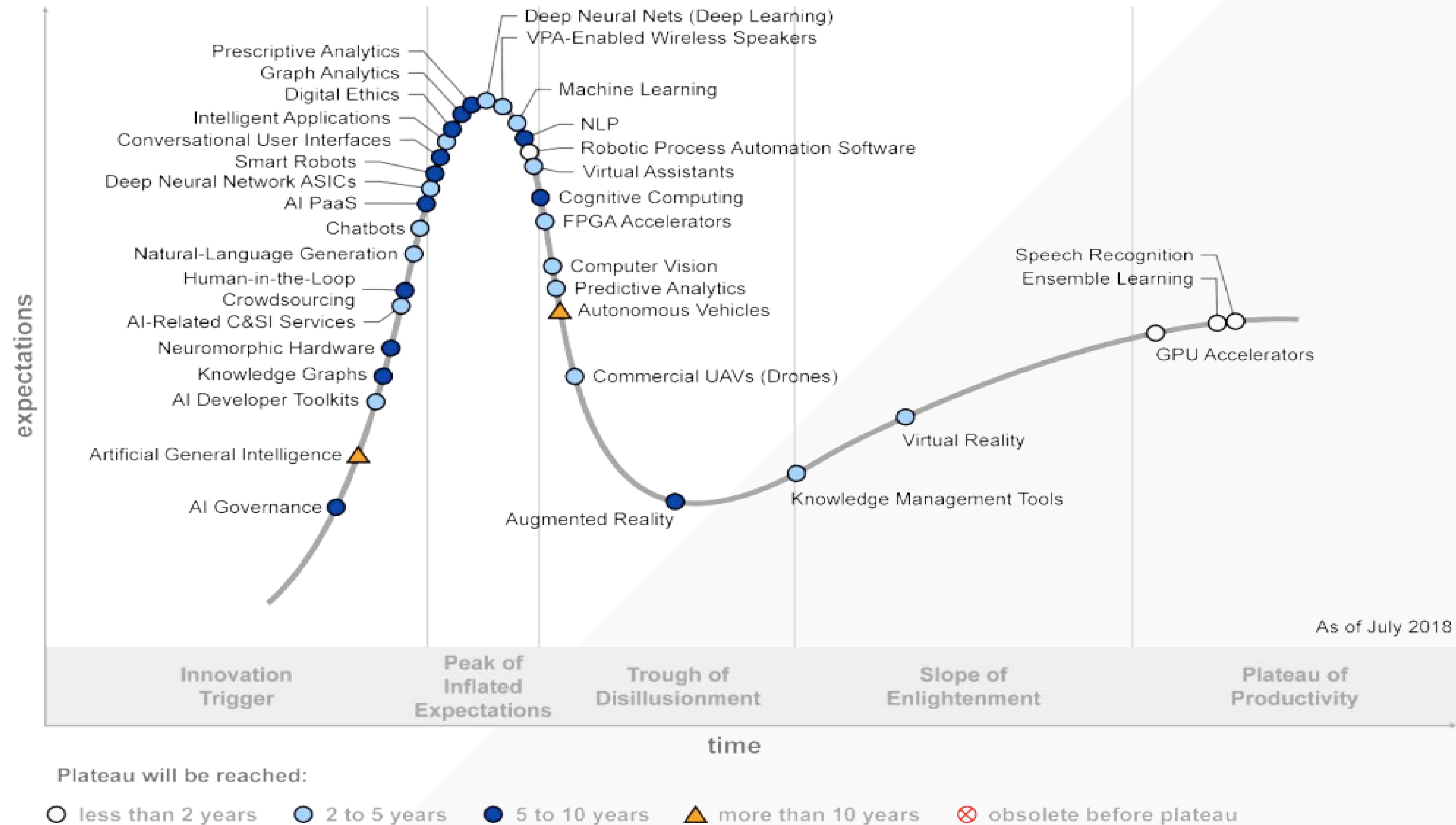
- 01 AI时代选择NLP不会错
- 02 入门NLP这些你要先知道
- 03 NLP的学习路线了解一下？
- 04 私人学习心得和资料推荐

PART ONE

01

AI时代选择NLP不会错

人工智能技术成熟曲线



Source: Gartner (July 2018)

© 2018 Gartner, Inc.

自然语言处理领域的典型企业

自然语言处理领域的代表性公司

排名	企业	综述
1	微软亚洲研究院	近些年在人工智能领域，尤其是自然语义理解、计算机视觉等方面取得了不菲的成绩。它是人工智能领域的“黄浦军校”。
2	搜狗	利用数据挖掘、自然语言处理技术实现搜狗输入法。
3	科大讯飞	一家专业从事智能语音及语音技术研究、软件及芯片产品开发、语音信息服务的国家级骨干软件企业。
4	腾讯	2017 年在西雅图成立人工智能实验室，其人工智能实验室副主任俞栋博士是前微软研究院语音识别专家。
5	华为	2016 年在香港成立诺亚方舟人工智能实验室，该实验室有一个部门专注于以无缝的方式和自然语言让机器与人沟通。
6	阿里巴巴	2018 年，阿里巴巴达摩院机器智能实验室开源了新一代语音识别模型 DFSMN，将全球语音识别准确率纪录提高至 96.04%。
7	云知声	语音识别领域的初创公司，在平台及行业领域，云知声已经与科大讯飞构成了直接竞争。
8	思必驰	思必驰对话工场是提供语音识别、语音合成、语义理解、智能对话以及声纹识别服务的开放平台。
9	出门问问	出门问问语音助手 APP 是一款基于语音识别、语义理解技术，面向移动搜索领域的智能手机应用。用户只需通过语音用口语化的提问方式就可以搜索出有关于吃穿住行的服务。
10	达观数据	国内唯一一家将自动语义分析技术应用于企业数据化运营的人工智能公司。

薪资待遇

自然语言处理隶属于人工智能，其职位需求和薪资待遇远高于其他岗位

智联招聘
zhaopin.com

职位 ▼ nlp算法工程师 🔍

自然语言处理（NLP）算法工程师(000957)

中移(杭州)信息技术有限公司 ZP名企

20K-30K 北京 1-3年 本科

年底双薪

交通补助

餐补

带薪年假

补充医疗保险

☆ 收藏

申请职位

自然语言处理（NLP）算法工程师(000957)

中移(杭州)信息技术有限公司 ZP名企

20K-30K 杭州 1-3年 本科

年底双薪

交通补助

餐补

带薪年假

补充医疗保险

国企

500-999人

招聘中

nlp算法工程师-杭州（恒生研究院）

恒生电子股份有限公司 最佳雇主

20K-40K 杭州 5-10年 硕士

五险一金

餐补

交通补助

弹性工作

节日福利

上市公司

1000-9999人

最新

nlp算法工程师(J180926002)

滴滴出行 ZP名企

20K-30K 北京 3-5年 本科

五险一金

年底双薪

绩效奖金

定期体检

带薪年假

外商独资

1000-9999人

招聘中

NLP相关岗位薪资

职位信息

岗位职责:

研究开放最新的机器学习算法并用于解决实际问题

构建NLP系统提升数据处理能力，进行实验，并改进产品

具有创造性思维能力，可以发现并解决新问题，从而发表论文或申请专利

任职资格:

自然语言处理，语音处理，机器学习，或相关领域博士或硕士学位

坚实的理论基础和NLP相关项目的实际动手能力

熟练掌握 Python 或 C++ 等编程语言

熟悉常用深度学习平台Tensorflow，Pytorch，或Caffe等

【滴滴出行】自然语言处理算法工程师其岗位职责，侧重于研究前沿技术，需要动手能力较强。

智能服务事业部自然语言处理算法工程师

面议

+

-

×

÷

阿里巴巴集团

2P名企

杭州 | 3-5年 | 本科 | 招1人

职位信息

职位描述

随着互联网的快速发展，自然语言的智能交互越来越普遍，搜索也正在从keyword搜索跨越到自然语言的语义搜索。该职位将负责自然语言的基础研究和开发（NLP）、自然语言的语义理解，打造阿里集团领先的智能人机交互系统，服务于全球用户。

1.负责商业智能领域语义分析相关技术和方法的研究；

2.负责利用自然语言处理技术和语义分析方法实现语义模型的建立与实践；

3.从系统应用的角度,利用语义分析的理论和方法解决实际问题。

1.扎实的算法基础：熟悉算法相关理论，算法原理和机器学习基本理论，具备扎实良好的数学基础；

2.精通自然语言处理相关领域的知识与技能；

3.在人工智能领域有一定的积累，能够从复杂的业务场景中进行算法选型、算法优化以及不断提升效果，并且能够真正应用在商业环境，成为商业智能的驱动者；

4.java/C++/Python基础扎实，熟悉hadoop、spark框架者尤佳；

5.在人机交互、智能问答、文本检索、多语言处理等等领域有实际的开发和从业经验者优先；

6.良好的团队合作意识，对技术抱有热情；

【阿里巴巴】招聘自然语言处理高级工程师，侧重算法理论基础和优化。这里的要求就相对滴滴出行需要更加坚实的理论基础。如算法基本理论，算法优化等。

自然语言处理有什么作用

- 机器翻译, 文字转语音, 翻译电话/电话转译...
- 双语词典自动建构, 双语词/短语自动对齐...
- 知识本体自动建构, 特定领域专业辞典建构...
- 论文自动写作, 拼字校正, 文法校正...
- 信息检索, 信息抽取, 数据挖掘, 文本挖掘, 网络挖掘...
- 文件/网页分类/分级, 垃圾信分类...
- 跨语言信息检索, 跨语言信息抽取, 跨语言...
- 生物信息, 序列比对, 结构比对...
- 情绪分析, 民意分析, 情报分析, 关键词拦截...
- 自动营销...
- 所有你想象得到, 跟 "文字" 或 "类文字" 的符号系列 (如 DNA, 乐谱) 相关的自动化分析, 转换, 合成, 要想做得更好, 最终都不能不考虑他们的结构跟功能 (语法跟语意). 所以, 自然语言处理的技术, 最终都会被用到这些应用领域.

“

那你说, 重不重要?

”

PART TWO

02

入门NLP这些你要先知道

自然语言处理基本概念

1、什么是“自然语言”(NL : Natural Languages)

- 自然语言就是我们一般人平常所用的, 所写的中文, 英文等语言文字 (广义来说, 也包含语音).
- 所谓的 "自然" 是相对于 "人工" 或 "程序" 语言 (Artificial or Programming Languages) 而言.
- 讲妈妈教的话, 或写文章, 就不一定要那么规规矩矩, 偶而出现火星文也很酷啊!! 不过, 酷归酷, 这种特性也让自然语言处理变得很难.

2、什么是“自然语言处理”(What is Natural Language Processing / NLP)

- 就处理的型态来讲, 大概可分为三类: 分析, 转换及生成.
- 分析(Analysis): 就是通过统计模型, 推理, 模式识别/分类, 机器学习等特异功能, 去 "分析" 文字的内涵, 结构, 了解文件在讲什么.
- 转换 (Transfer): 分析之后, 可以将他 "转换" 成另一种形式的有用信息, 作进一步的应用. 比如转到另一种语言的深层结构 (自动翻译), 或数据库 (数据仓储).
- 生成 (Generation): 有时候, 我们也会把某些有用, 但是比较抽象的信息, 用文字写出来, 或用语音说出来, 这叫 "生成" 或 "合成".

如何将自然语言转化成计算机语言？

我	0	0	1
和	0	1	0
你	1	0	0

稀疏向量

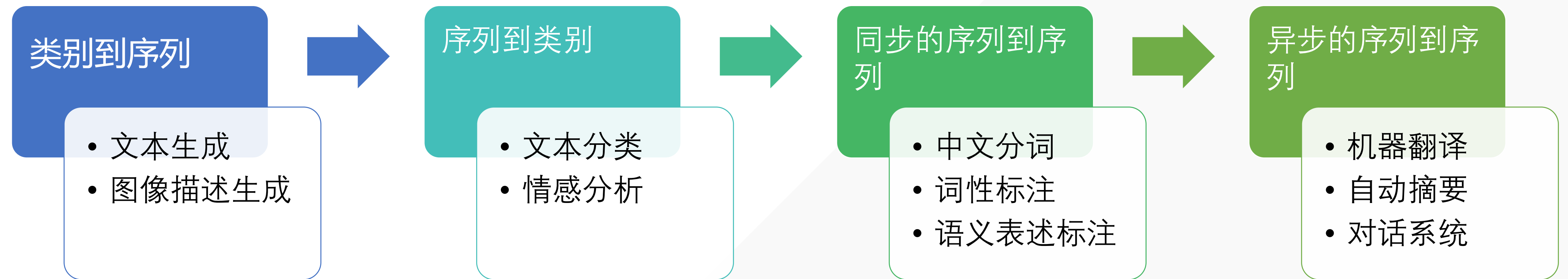
One-hot

我	0.2	0.8	1.3
和	0.3	1.9	0.1
你	0.2	0.5	1.1

稠密向量

- 1.word2vec、glove
- 2.ELMO、GPT
- 3.BERT

NLP的应用类型




NLP的实际应用：舆情监测

微博热点事件



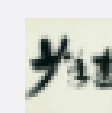
澎湃新闻

1月7日 22:01 来自 微博 weibo.com 已编辑

【“糖丸之父”当年携幼子以身试药，我们都欠他一声“谢谢”】#送别糖丸之父顾方舟#  央视新闻的秒拍视频 1月2日，我国病毒学家顾方舟逝世，享年92岁。昨天，中国医学科学院为他举行了追思会。他受命于危难之时，曾拿自己刚满月的儿子做试验，于是，他不到1岁的大儿子成了中国首个喝下疫苗的孩子。他说：“我们如果不相信，别人怎么能相信。”他发明的“糖丸”消灭了中国的小儿麻痹症，使数十万儿童免于致残。@央视新闻

+关注

按热度 | 按时间

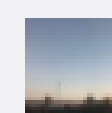


1少年游1：老爷子看到现在的假疫苗该有多痛心

1月7日 22:02

回复 | 3123

还就这样了 等人 共22条回复



祈雨月：谢谢您将药做的这么好吃！❤️

1月7日 22:02

回复 | 1418

moriO_O 等人 共16条回复

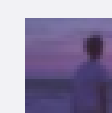


律策律所王成岩：当年“携幼子以身试药”，当今“拿别人孩子修改基因”，这就是社会的变化，把人性、伦理、道德、规则都变得若有若无了

1月7日 22:08

回复 | 995

Frankwung1 等人 共17条回复

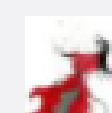


猫巷里情话：一路走好致敬科学家

1月7日 22:03

投诉 | 屏蔽 | 回复 | 369

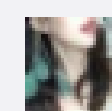
保持单身的男人 等人 共1条回复



阿腾LYZ：谢谢您 谢谢您的糖丸儿

1月7日 22:02

回复 | 318



迷妹喵喵：有的人就是来拯救世界的！伟大！

1月7日 22:18

回复 | 242

NLP在搜索引擎中的应用

◆传统的搜索引擎：

根据关键词匹配用户想要的搜索的信息，无法真正理解用户的意图，使得用户不得不一次又一次的调整关键词，才能得到更贴近自己需要的信息。

◆智能搜索引擎：

（1）为了使搜索引擎更为准确的理解用户输入的自然语言信息，我们必须要对用户输入的句子进行分解，也就是把用户输入的语言单元化，我们也叫做分词，它对搜索引擎的影响是非常大的，划分的单元过小，会占用大量空间，影响查询速度，划分的单元过大，又会影响查询的准确度。

（2）理解用户所搜索的关键词的含义及多个关键词之间的关联，从而找出用户真正需要的结果。

NLP的实际应用：推荐系统

个性化推荐是大数据时代不可或缺的技术，在电商、信息分发、计算广告、互联网金融等领域都起着重要的作用。

在推荐系统中经常需要处理各种文本类数据，例如商品描述、新闻资讯、用户留言等等。

刚过去的双十一，双十二，在淘宝，京东上选购产品，平台会非常快速的推荐人们更倾心的产品。A

- 候选商品召回：候选商品召回是推荐流程的第一步，用来生成待推荐的物品集合。
- 相关性计算：文本相似度算法以及用户画像计算时用到的一些相关性计算



NLP的实际应用：机器翻译

谷歌翻译，2017 年宣布其机器翻译实现了完全基于 attention 的 transformer 机器翻译网络架构，实现了新的最佳水平。

Google 翻译

为什么“机器翻译”是明日之星

- 网络打破了距离的障碍,
- 但是... 语言障碍 (Language Barrier) 仍旧存在！
- 语言成了目前妨碍信息流通及存取的最大绊脚石.
- 跨语言的知识抽取与分享, 必须有强力的“翻译引擎”才容易达成目标.

有道翻译 免费、即时的多语种在线翻译

PART THREE

03

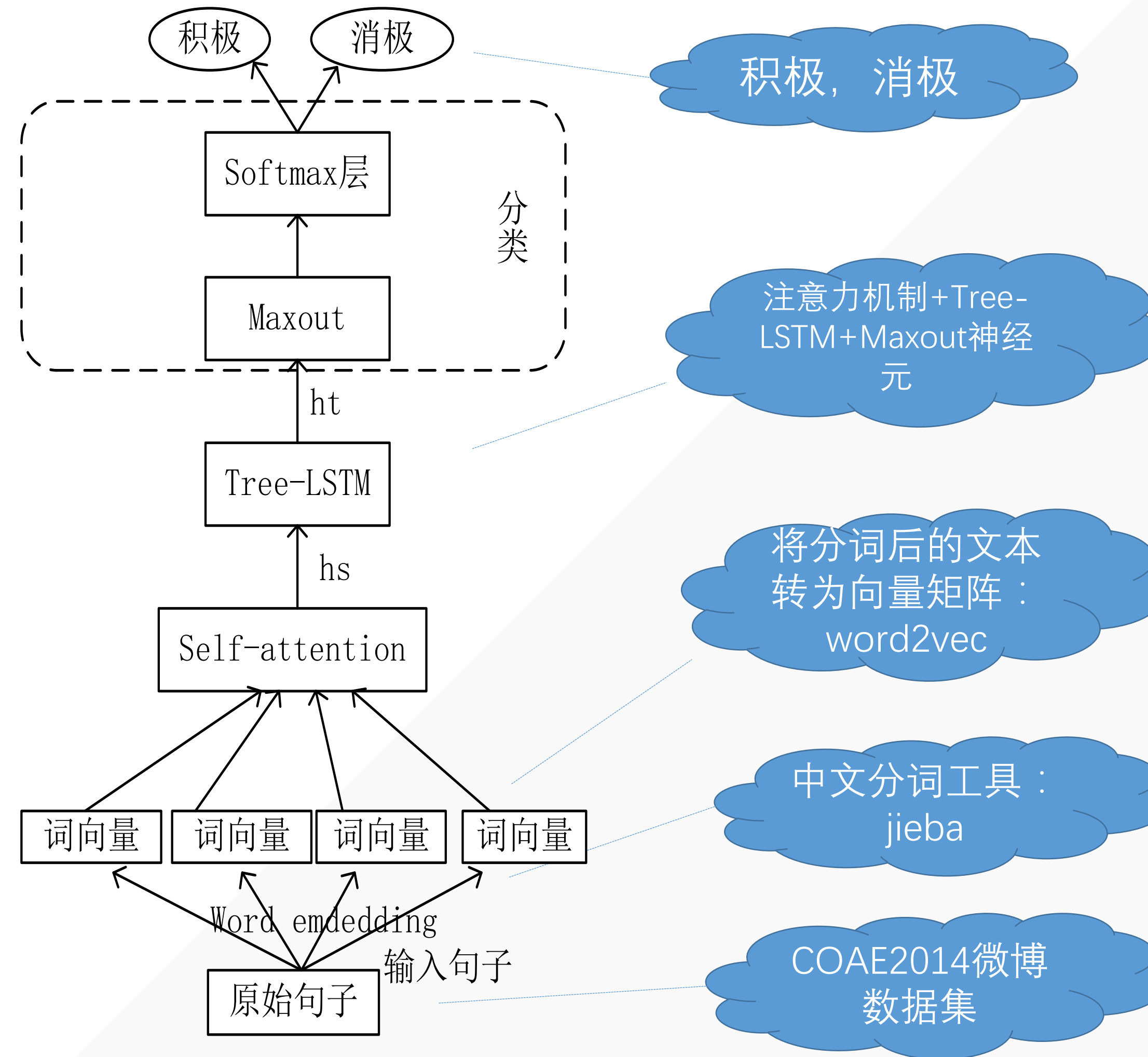
NLP的学习路线了解一下？

自然语言处理基本知识学习路线

自然语言处理学习要有计划，切勿乱学。



情感分析研究框架



自然语言处理常用的算法



逻辑回归 (logistic) ,
支持向量机 (SVM)



卷积神经网络 (CNN)



注意力模型 (attention)



主题模型LDA



采用td-idf, text-rank



序列模型 (seq2seq)

PART four

04

私人学习心得和资料推荐

关于学习的一些心得

关于学习

- 先学一个深度学习框架
- 从小项目开始小实战（写过的代码注明功能）

关于搜索

- 多搜github
- 多看博客（自己试着写博客，收获很大）
- 关注NLP相关的公众号

关于比赛

- kaggle
- 天池
- AI challenger

学习平台

- google提供的官方文档
- google免费的GPU Colab

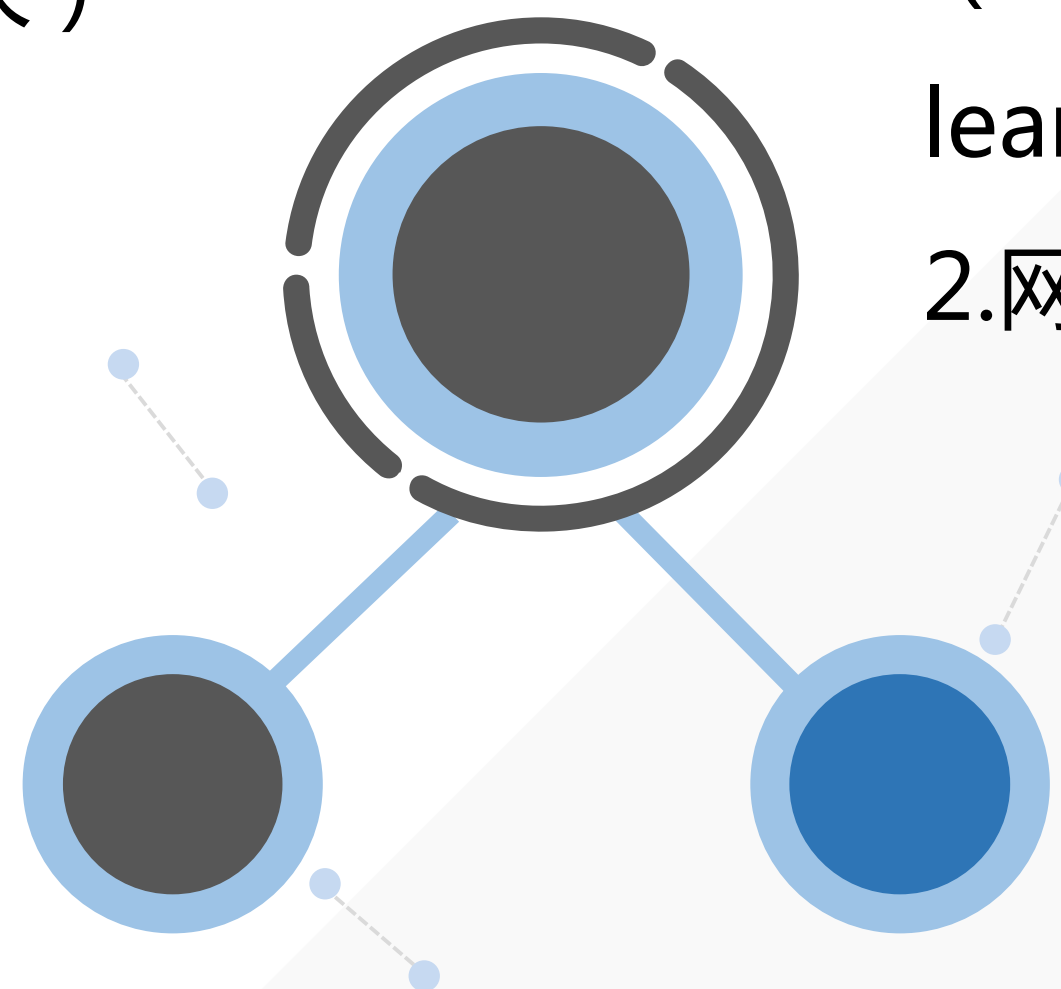
资料推荐

书籍

1. 吴军老师的《数学之美》
2. 《统计自然语言处理(第2版)》（宗成庆）蓝皮版
3. 《西瓜书》（周志华）
4. 《统计学习方法》（李航）
5. 《自然语言处理综论》（斯坦福大学 Dan Jurafsky 教授和科罗拉多大学的 James H. Martin 教授）

课程

1. google出版的官方机器学习速成课程
(<https://developers.google.cn/machine-learning/crash-course/>)
2. 网易云自然语言课程



其他资源

1. 我爱自然语言处理
2. 苏剑林的科学空间
3. paperweekly

AI工程师（自然语言处理）微专业

⑤ 微专业 / AI工程师(自然语言处理)

从入门到精通，解锁NLP工程师必备技能

- 35个实战案例，9大应用场景，足量GPU在线实验平台，机器学习&深度学习双套解决方案，带你成为Offer收割机 -

¥2698 原价：¥3800

¥200 01月17日前有效

立即购买

Ø 关键的NLP核心知识

Ø 6大NLP细分领域方向

Ø 35个前沿技术实战案例

Ø 专业的论文代码复现

Ø 足量的GPU实训环境

Ø 持续的岗位信息推送

欢迎大家和老师一起，加入AI工程师（自然语言处理）微专业：<https://url.163.com/mGk>

今天报名的同学可领取直播专属300元优惠券

学习交流、福利领取



扫码添加 我是 404 个人微信

拉你进网易云课堂AI交流微信群

微信号：**netease404**

干货学习资料、免费体验课

领取NLP课程**直播专属300元**优惠券

Thank you