自然语言处理如何入门

--NLPer的打怪升级之路

张奇

复旦大学

0 NLP的研究早在第一台电子计算机出现后不久就已经开展

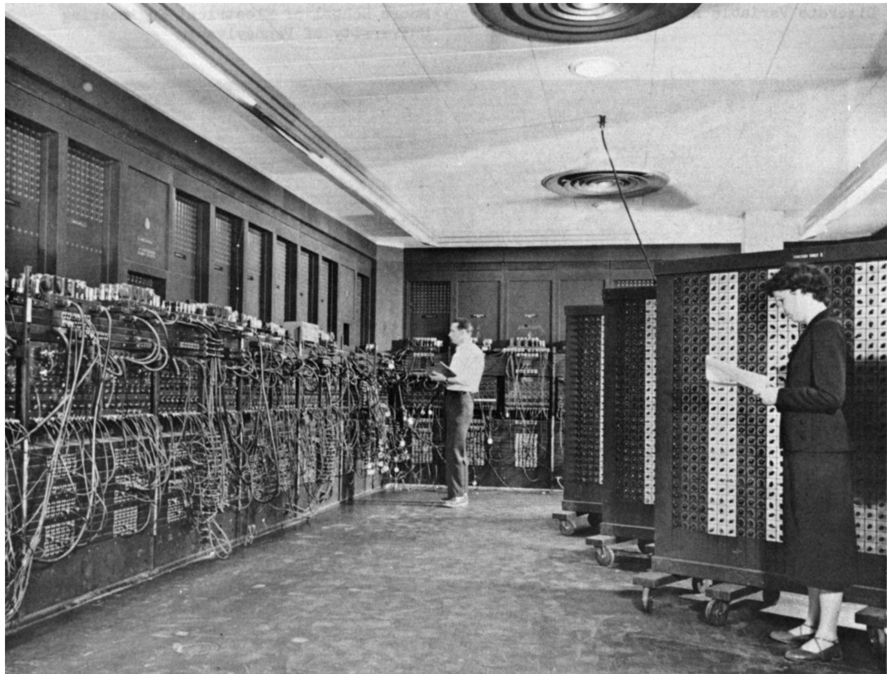
1946年2月14日，由美国军方定制的世界上第一台电子计算机“电子数字积分计算机”（ENIAC Electronic Numerical And Calculator）

1950年，图灵发表论文《计算机器与智能》

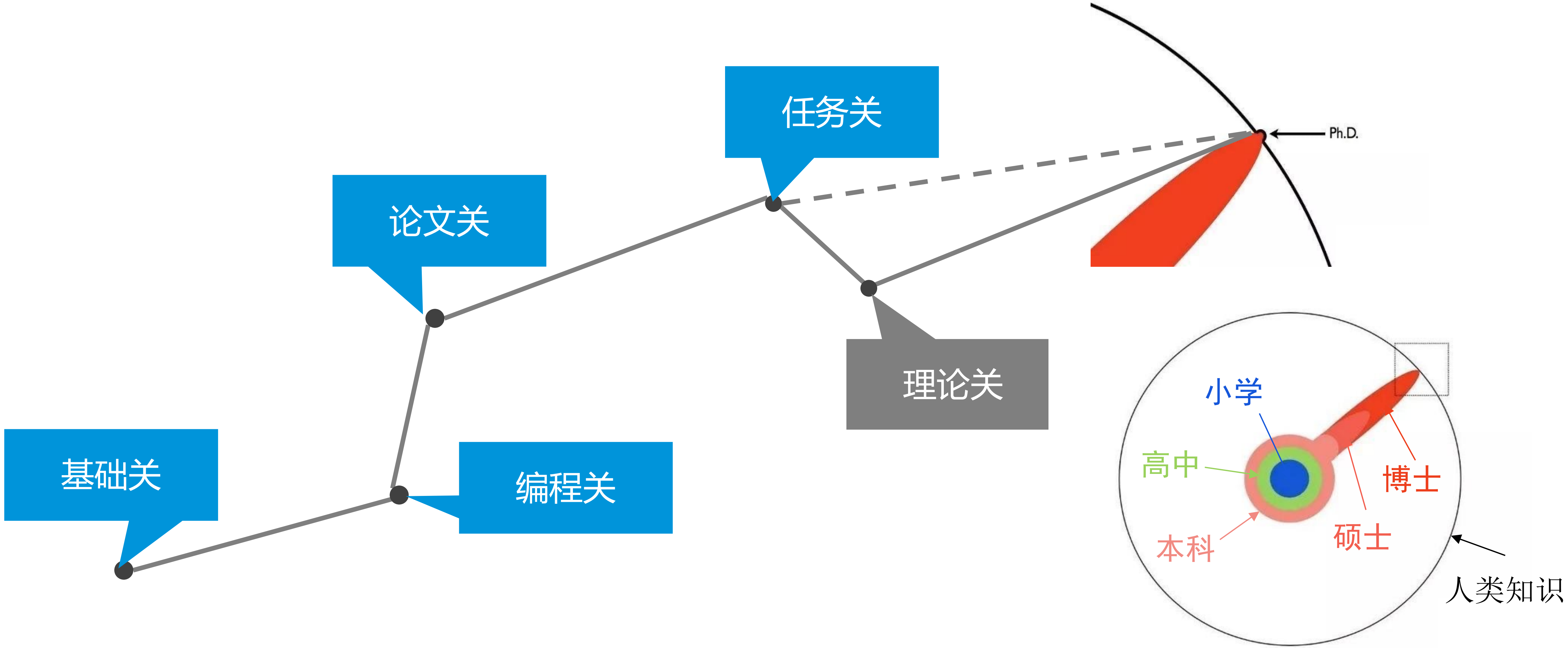
1954年，IBM实验室进行了自动翻译实验，将60句俄文自动翻译为英文

Natural Language Understanding

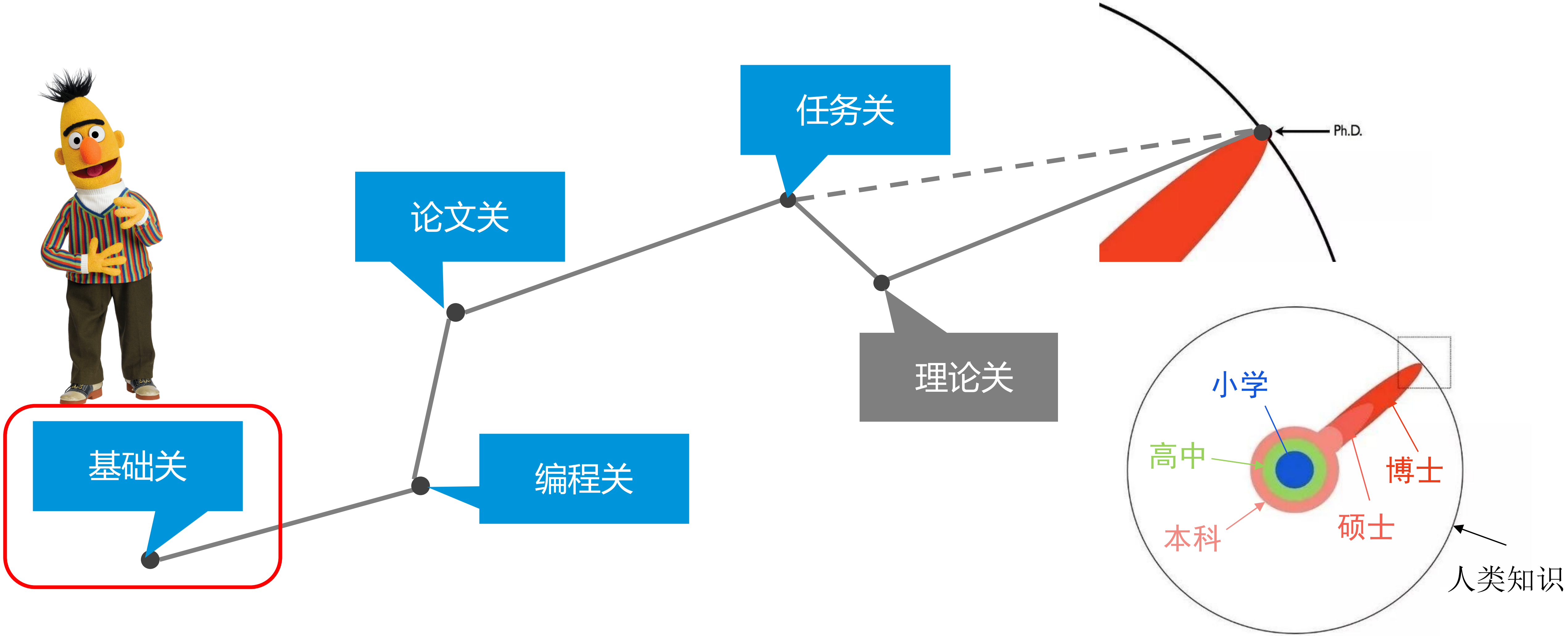
Natural Language Processing



0 NLPer的打怪升级路线图



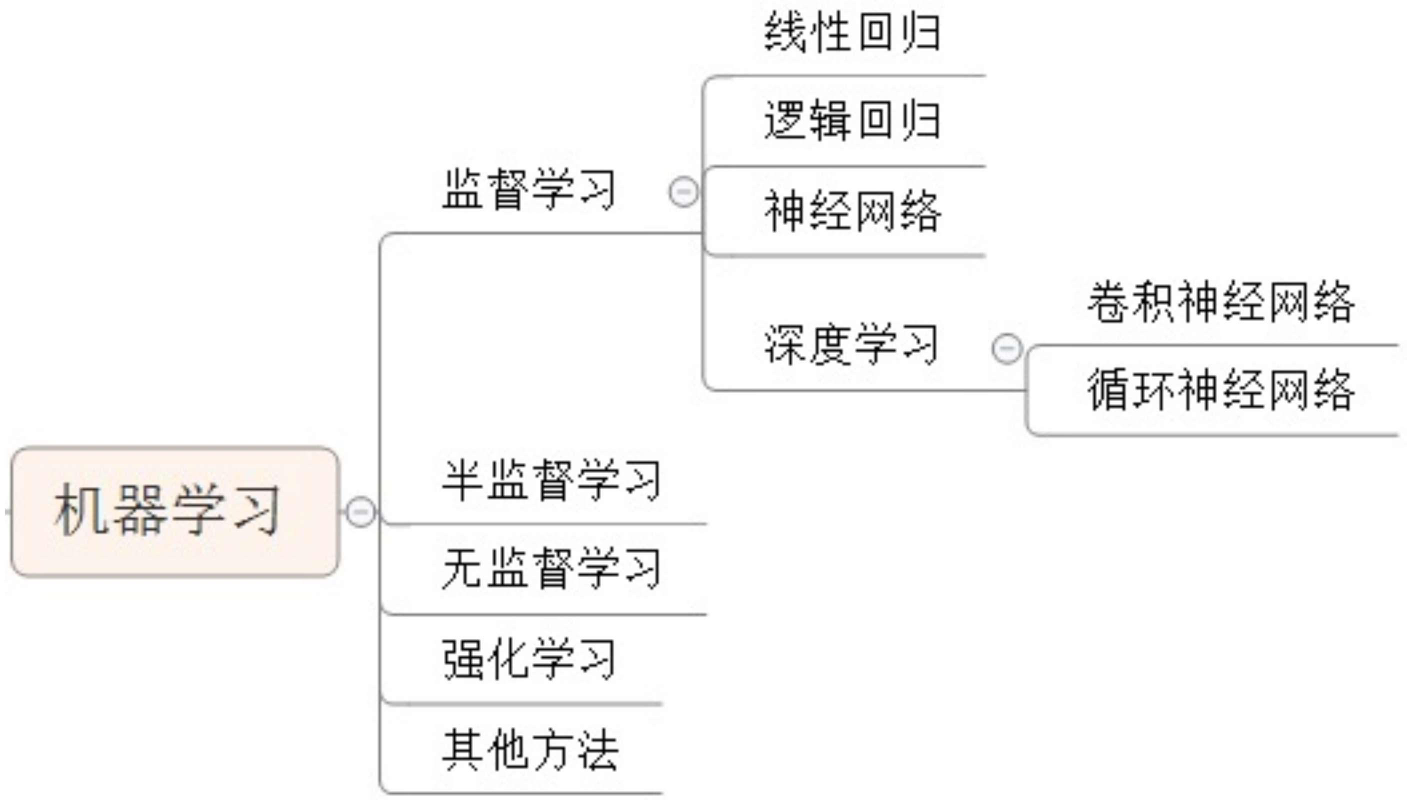
0 NLPer的打怪升级路线图



1 基础关—目标1：NLP基本任务、经典理论

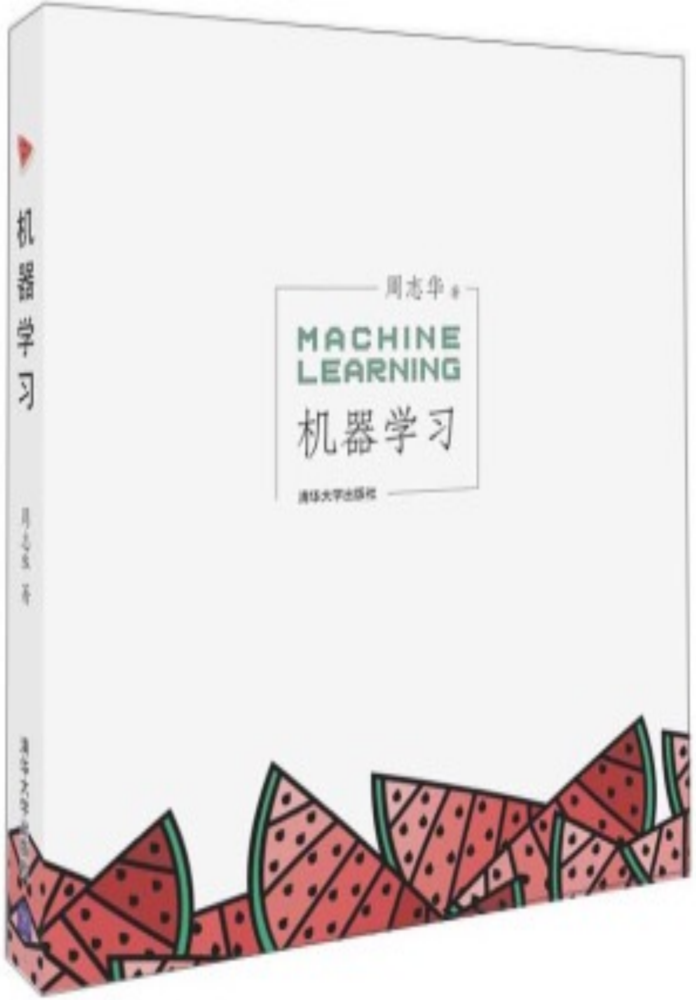
- 类别到序列
  - 文本生成
  - 图像描述生成
- 序列到类别
  - 文本挖掘
  - 文本分类
  - 情感分析
- 同步的序列到序列
  - 中文分词
  - 词性标注
  - 实体识别
- 异步的序列到序列
  - 机器翻译
  - 自动摘要
  - 问答系统

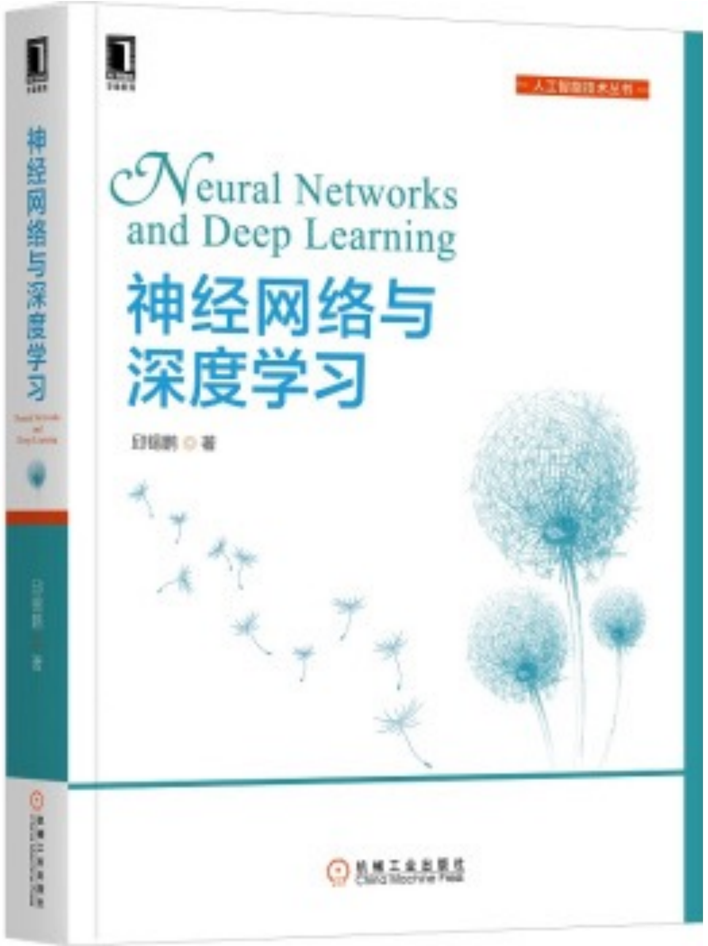
1 基础关—目标2：机器学习基本任务、经典理论

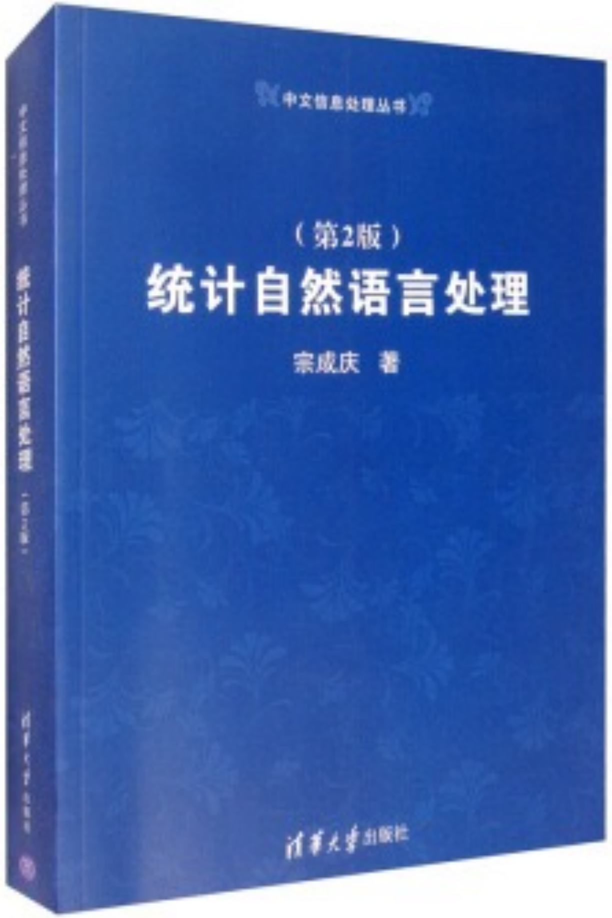


1 基础关--如何过关

基础


机器学习
周志华 著
MACHINE LEARNING
机器学习

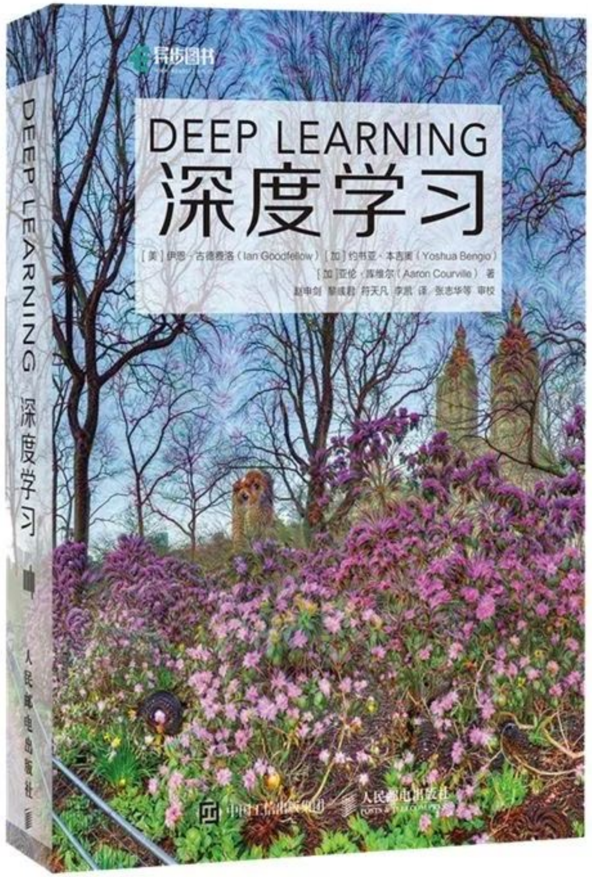

Neural Networks and Deep Learning
神经网络与深度学习
神经网络与深度学习


（第2版）
统计自然语言处理
宗成庆 著
清华大学出版社

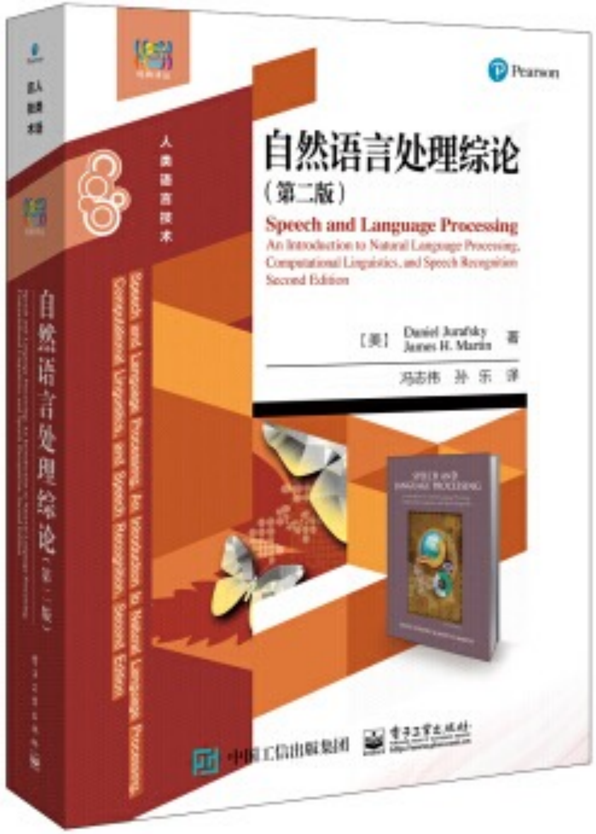
1 基础关--如何过关

提高

统计学习方法
(第2版)
李航 著
清华大学出版社



DEEP LEARNING
深度学习



Pearson
自然语言处理综论
(第二版)
Speech and Language Processing
An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition
Second Edition
[美] Daniel Jurafsky James H. Martin 著
冯志伟 孙乐 译

相关资源

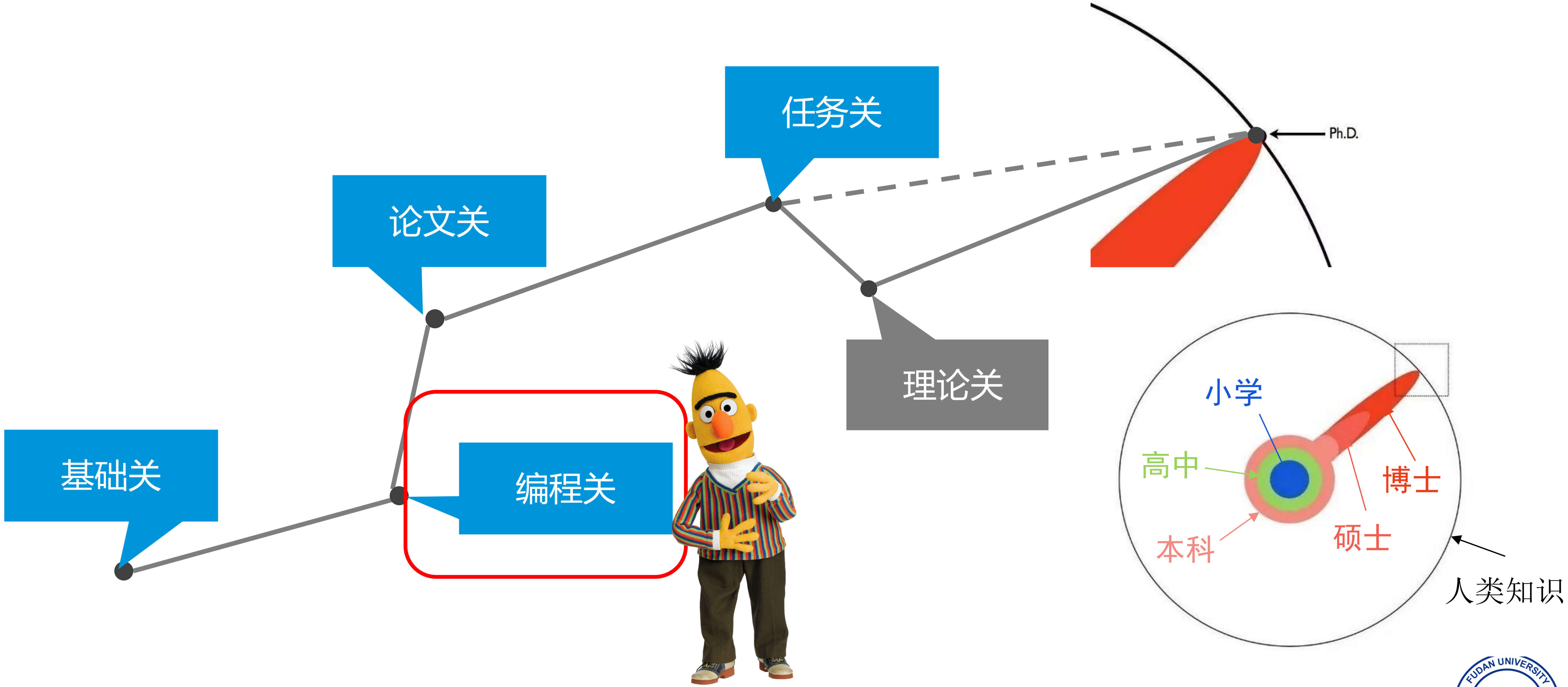
李宏毅2020机器学习深度学习(完整版)国语 https://www.bilibili.com/video/BV1JE411g7XF

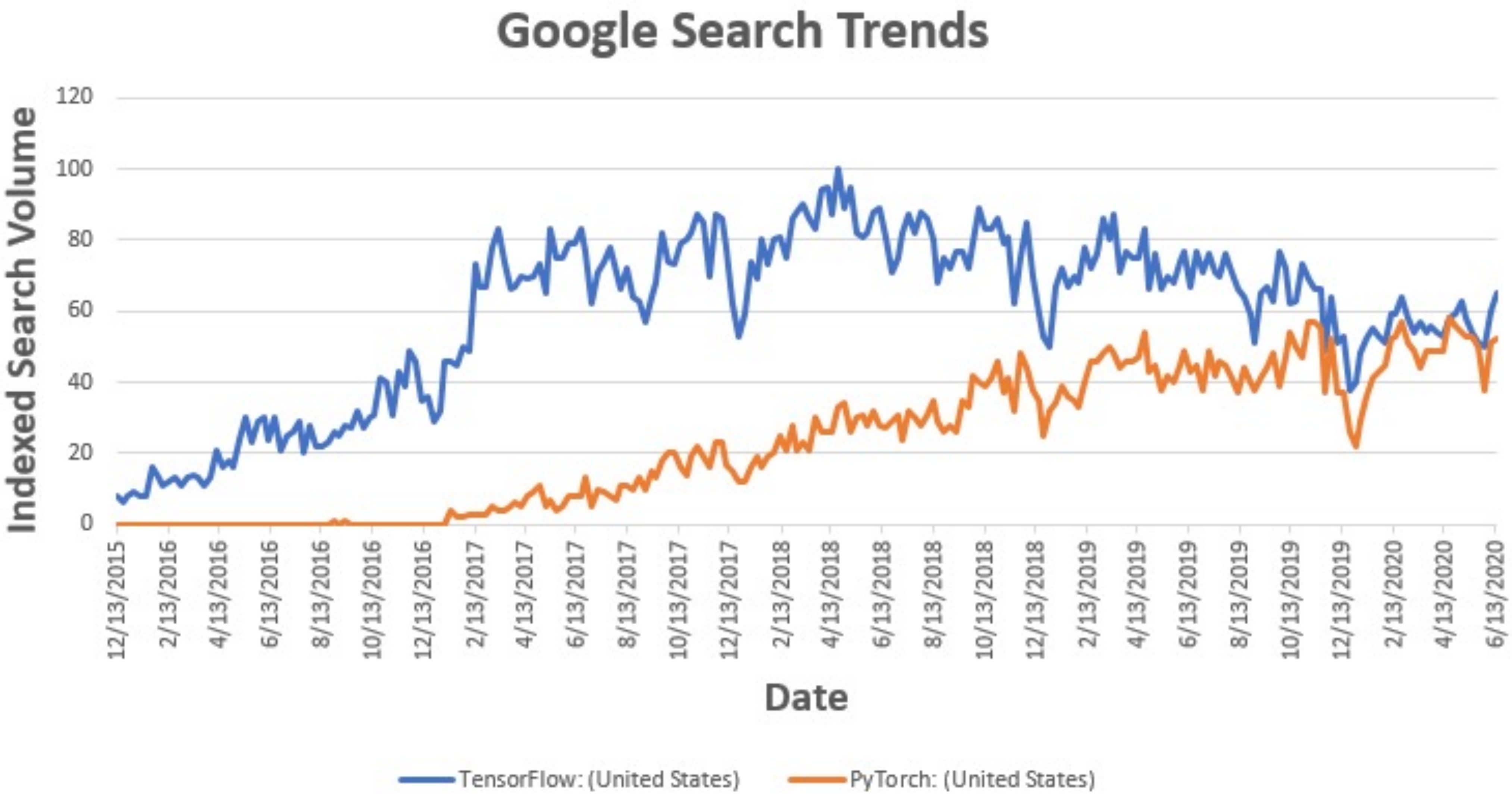
斯坦福CS224N《深度学习自然语言处理》课程(2021) by Christopher Manning https://www.bilibili.com/video/BV1nP4y1j7rZ

神经网络与深度学习 邱锡鹏 https://www.bilibili.com/video/BV13b4y1177W

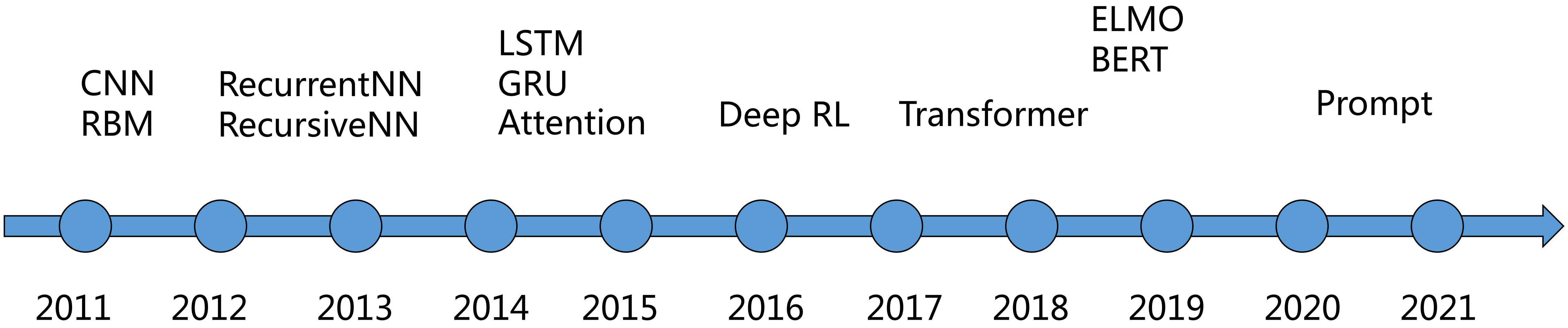
中科院 宗成庆 自然语言处理公开课 https://www.bilibili.com/video/BV1Cb411T7Cd

0 NLPer的打怪升级路线图

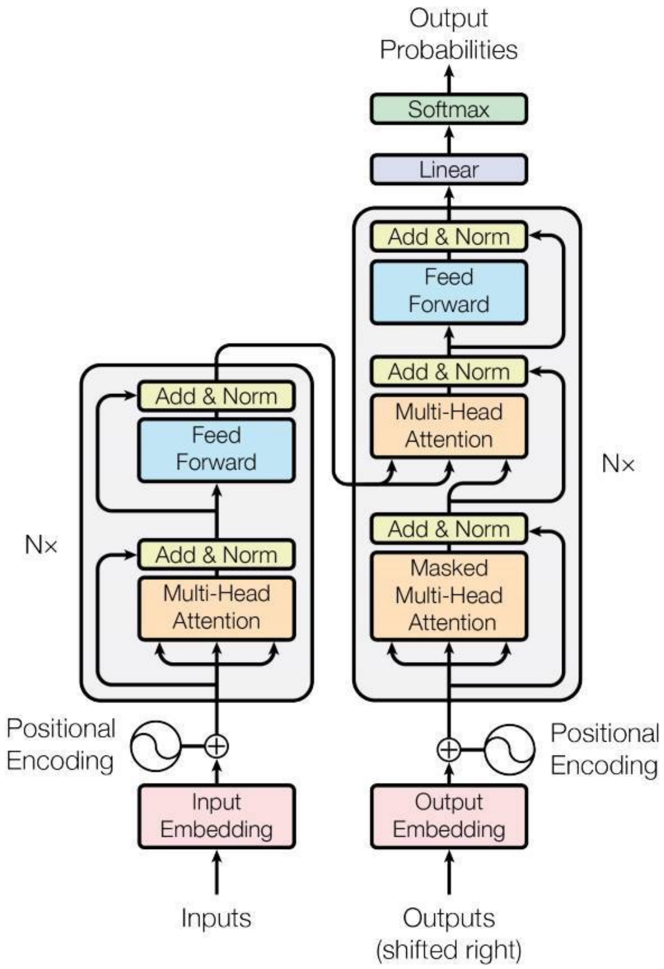


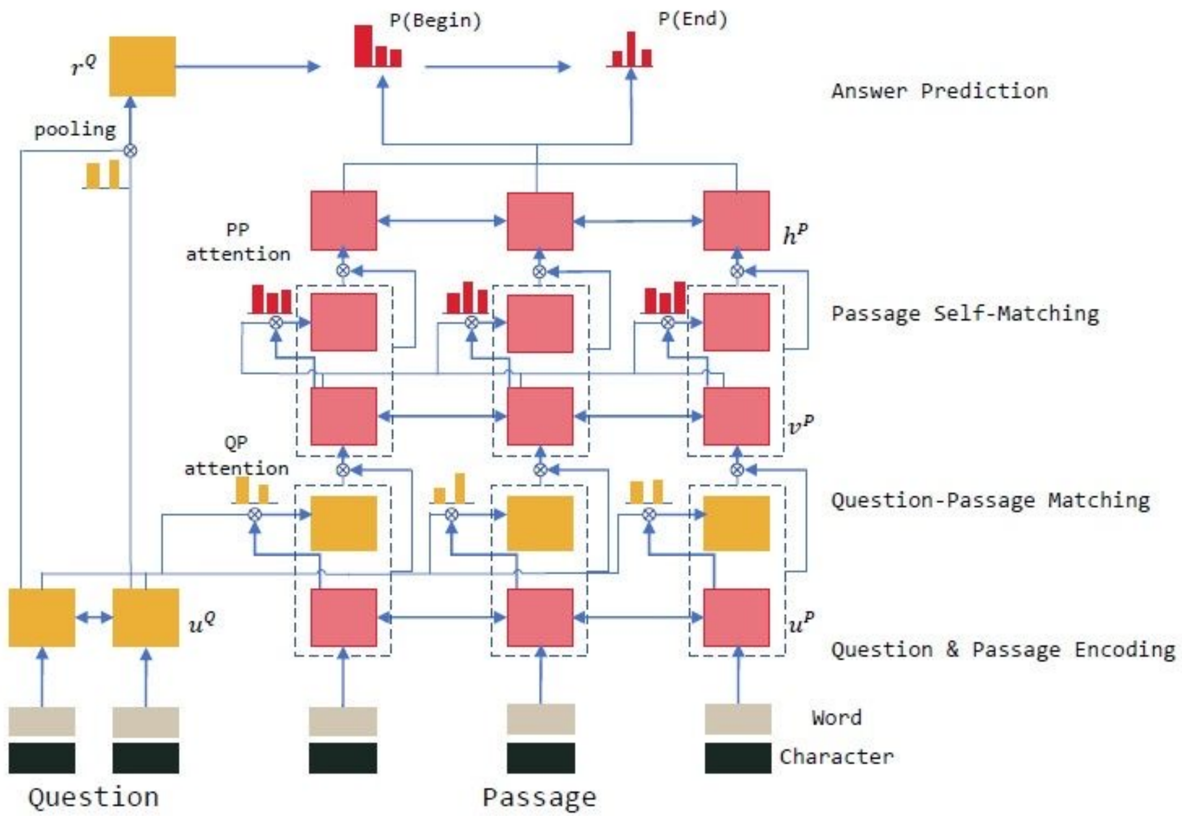

Google Search Trends
Indexed Search Volume
120
100
80
60
40
20
0
12/13/2015
2/13/2016
4/13/2016
6/13/2016
8/13/2016
10/13/2016
12/13/2016
2/13/2017
4/13/2017
6/13/2017
8/13/2017
10/13/2017
12/13/2017
2/13/2018
4/13/2018
6/13/2018
8/13/2018
10/13/2018
12/13/2018
2/13/2019
4/13/2019
6/13/2019
8/13/2019
10/13/2019
12/13/2019
2/13/2020
4/13/2020
6/13/2020
Date
TensorFlow: (United States)
PyTorch: (United States)

2 编程关—每年算法都有新的热点



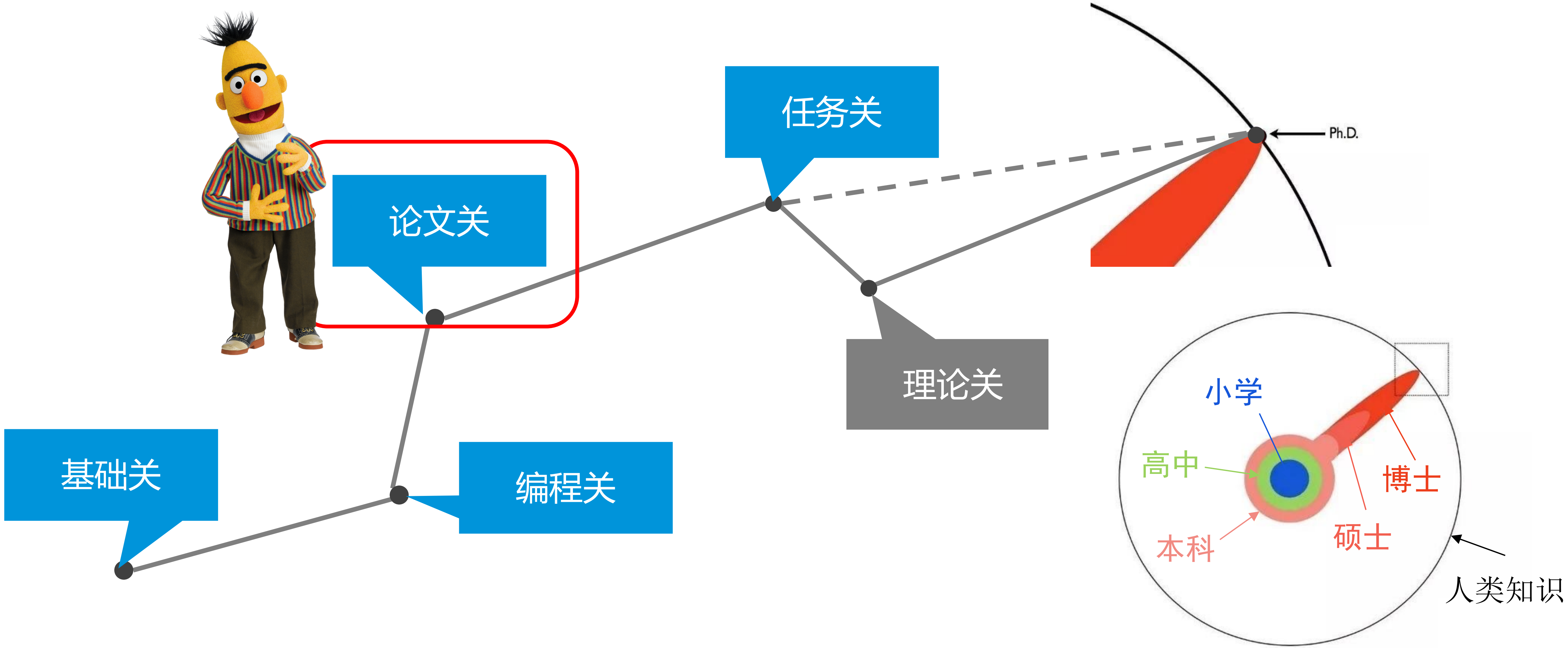
编程关—如何过关





1周内可独立实现R-NET 、Transformer基本结构

0 NLPer的打怪升级路线图



3 论文关—NLP&ML相关论文每年10K+

ACL 2021 3350篇论文投稿，最终有21.3%的论文被主会录用，并额外接收了14.9%的论文到Findings子刊。

EMNLP 2021共收到有效投稿3114篇，**录用**754篇，**录用**率仅为24.82%。

ICML 2021的**论文**接收结果已经公布，今年一共有5513篇有效投稿，其中1184篇**论文**被接收，接收率为21.5% 。

ICLR 2021 共收到了2997篇**论文**投稿，相比去年的2594篇**论文**投稿，增加了15.5%。 其中860篇**论文**被接收，接受率为28.7%

NeurIPS 2021 收到有效**论文**投稿9122 篇，接收率为26%，只有3% **论文**被接收为Spotlight。

3 论文关—分层过滤，集团作战，公众号热点

100+300 Paper Reading Group + Topic

500 Abstract、Introduction

1000+ Title

2000+ Best Paper、Oral、Spotlight、Session、重点单位

10000+ ACL、EMNLP、NAACL、COLING、IJCAI、AAAI
ICML、NIPS、ICLR、CVPR、ACM MM

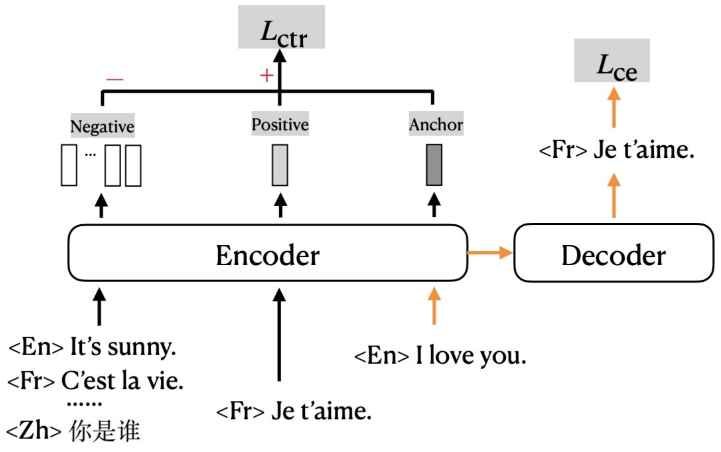
Contrastive Learning for Many-to-many Multilingual Neural Machine Translation

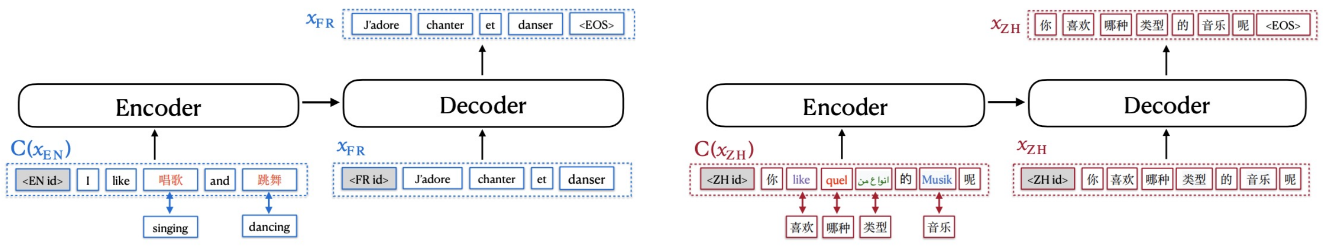
Motivation：当前机器翻译任务还是以英语为中心，导致非英语方向发展滞后，本文中作者想要提出一个many-to-many的跨语言通用翻译模型。

Methods：

1.用对比学习的方法将多种语言映射到同一个语义空ss-间中，令不同语言同语义的句子在表示空间中尽量接近。

2.对句子中的phrase进行aligned的跨语言替换，对平行语料和单独语料的数据增强方法如下图





(a) AA for Parallel Corpora (b) AA for Monolingual Corpora

3 论文关—如何过关

- **Making the Relation Matters: Relation of Relation Learning Network for Sentence Semantic Matching（周玮康）**

motivation：许多学者任务关系和关系之间也存在着信息，即“隐藏关系信息”。因此，不同重构之间的比较和对比学习（如成对关系学习）可以帮助模型更多地了解重构中隐含的语义信息，进而有助于增强模型的句子分析能力。它们应该被视为不仅仅是毫无意义的单热向量。

方法简介：首先利用预先训练过的BERT从全局的角度来建模输入的单词和句子的语义意义。然后，开发了一个基于cnn的编码器，从局部的角度获取句子的部分信息（关键词和短语信息）。接下来，受BERT训练处理中的自监督学习方法的启发，我们提出了一个关系的关系(R2)分类任务，以提高R2-Net对不同关系对应的内隐共同特征的学习能力。

此外，利用**三重损失**来约束模型，从而更好地分析类内和类间的关系。沿着这条线，具有相同关系的输入句子对将表示得更近，反之亦然则更远。将关系信息正确地集成到句子对建模处理中，有利于解决上述挑战，提高模型性能。

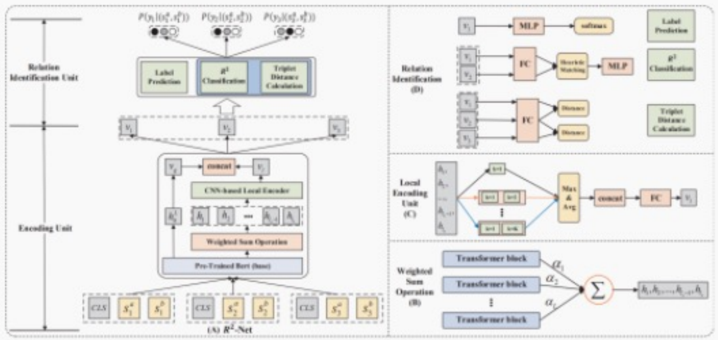


Figure 2: Architecture of *Relation of Relation Learning Network* (R^2*-Net*).

任务1：通过BERT全局编码，然后CNN局部编码，拼接得到v，最后MLP完成标签预测。

任务2：使用2个语对v1,v2，预测关系的关系。

任务3：从基于三重态的关系中学习类内和类间信息

最后将3个LOSS拼接，训练。**最后在Quora和MSRP上分别提升0.6%和0.1%**

ICML 2021: Weight-Covariance Alignment for Adversarially Robust Neural Networks

Motivation

1. 现有的stochastic neural networks (SNN) 都是启发式设计的,通常基于对抗训练,使得计算开销很大;
2. SNN注入噪声通常为isotropic noise,本文的SNN模型可以学习得到一个anisotropic noise分布来优化对抗鲁棒性的学习理论界限;

对抗防御方法存在的缺陷:

1. 使用干净数据和对抗数据的混合,而对抗样本的生成会增加计算复杂度和更长的训练过程;
2. 现有对抗防御方法都是经验启发的,缺少理论支撑;
3. 现有stochastic模型使用的noise都是isotropic,这意味着多元高斯分布拥有对角方差矩阵;这意味着扰动特征的时候,对于各个维度的扰动都是独立的;无相互依赖关系.

Stochastic Adversarial Defense

idea: random self ensemble: one can simulate an ensemble of virtually infinite models while only trainning one

inject fixed or learnable noise into the models

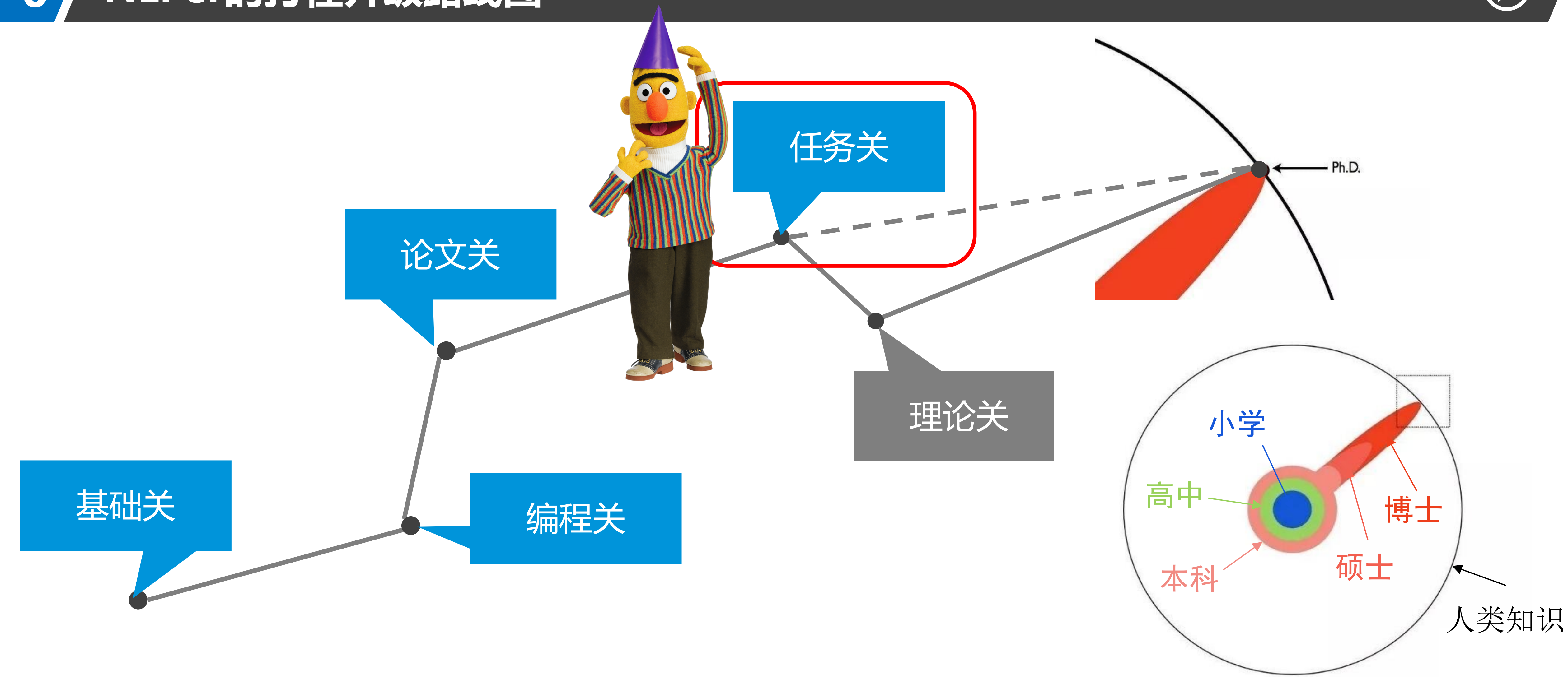
输入上的扰动,权重和表示上的扰动

Method

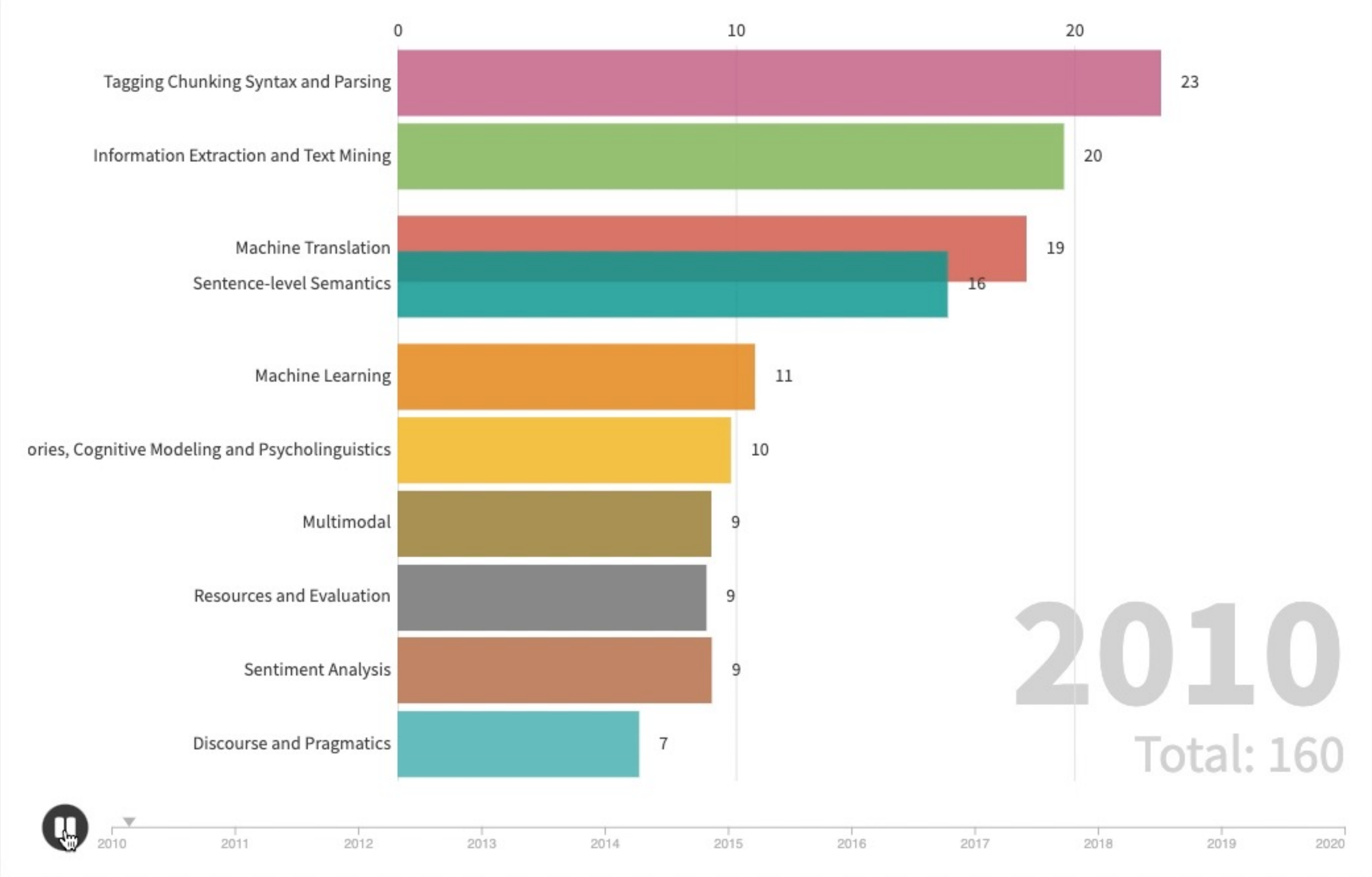
在分类层之前的表示上增加扰动,

Our method fits into the family of SNNs that apply additive noise to the penultimate activations of the network. Consider the function, $f(\vec{x})$, which implements the feature extractor portion of the network i.e., everything except the final classification layer. Our WCA-Net architecture is defined as

0 NLPer的打怪升级路线图



4 任务关—研究热点切换速度快

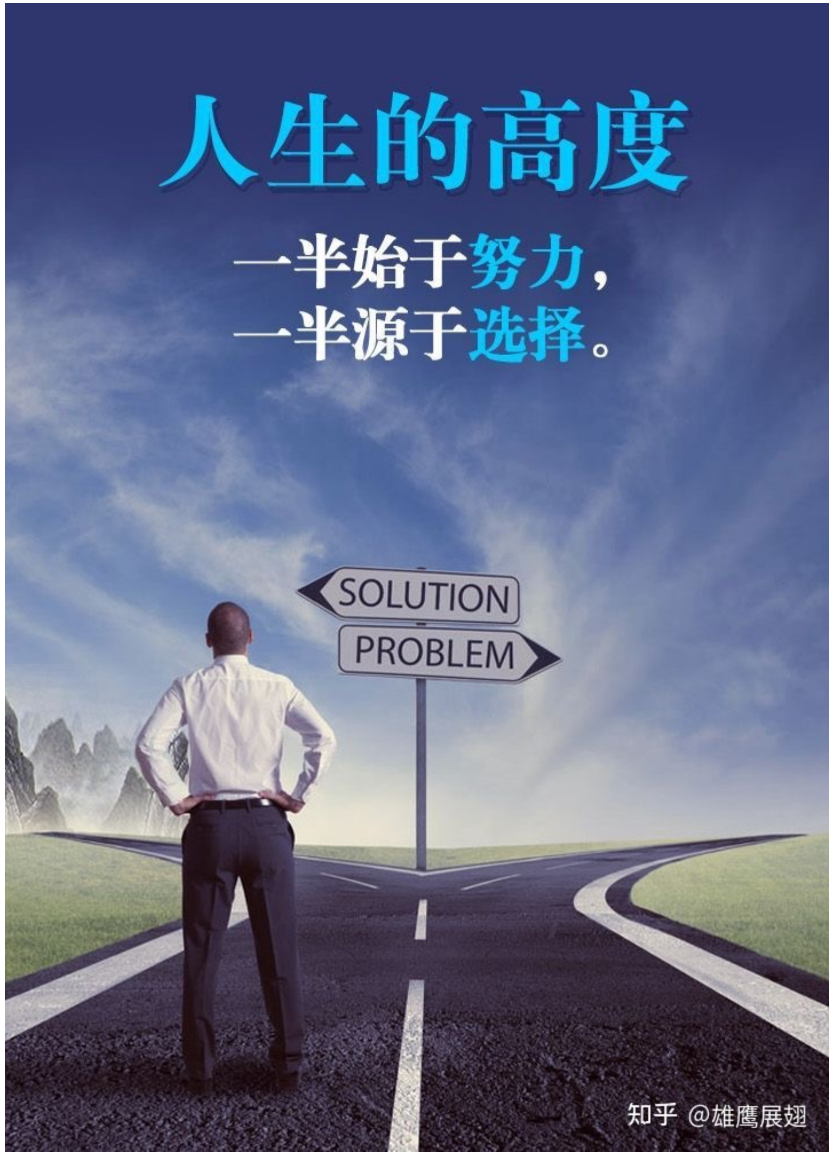


4 任务关—热门方向年年换，我该如何选

选择热门的方向

选择冷门的方向

4 任务关—热门方向年年换，我该如何选



1. 有明显的益处；

2. 描述简单；

3. 尚且不具备明确的解法；

4. 解决方案具备可测试性，大问题可以拆解为多个子问题，能观察到每个子问题的进程；

5. 要确保研究中数据获取的可靠性以及足够数量

《学术研究你的成功之道》--- 凌晓峰，杨强

4 任务关—热门方向年年换，我该如何选

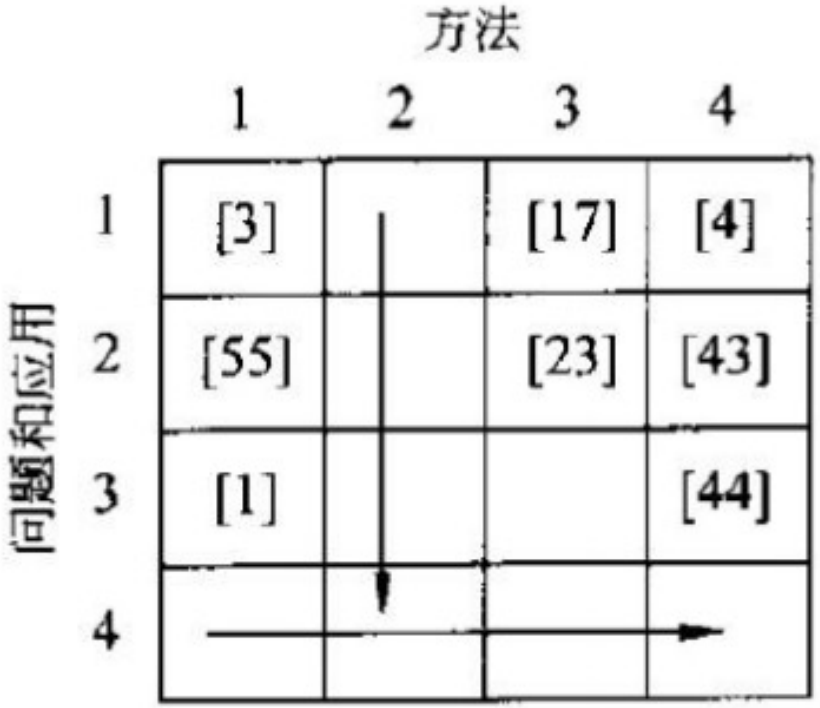


图 4.1 棋盘法举例

- **横轴**代表相应的研究领域用到的不同方法、方案和技术
- **纵轴**表示要解决的研究问题或潜在的研究问题
- 具有相关性的研究问题要按照顺序排放，无相关性的研究问题可以随意摆放

《学术研究你的成功之道》--- 凌晓峰，杨强

4 任务关—热门方向年年换，我该如何选

	资讯	交流	娱乐	商务
Web1.0 Search	Baidu百度		SNDA	阿里巴巴
Web2.0 Social				?
Web3.0 Mobile	?	?	?	?

“四纵三横”理论---王兴

4 任务关—什么样的Idea可以发论文

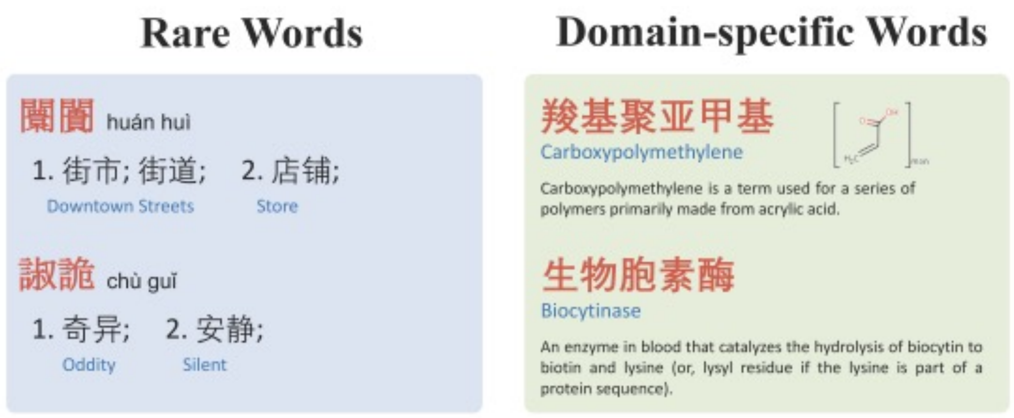


Figure 1: Examples of rare words and domain-specific words in dictionaries.

Figure 1: An example of multimodal tweets. In this tweet, "*Rocky*" is the name of the dog.

之前的分词算法在处理未登录词或者处理领域特有词时，会有问题

之前的NER算法不会关注图像信息，但是没有图像信息的话，一些Tweets不能很好分析

4 任务关—如何找到Motivation

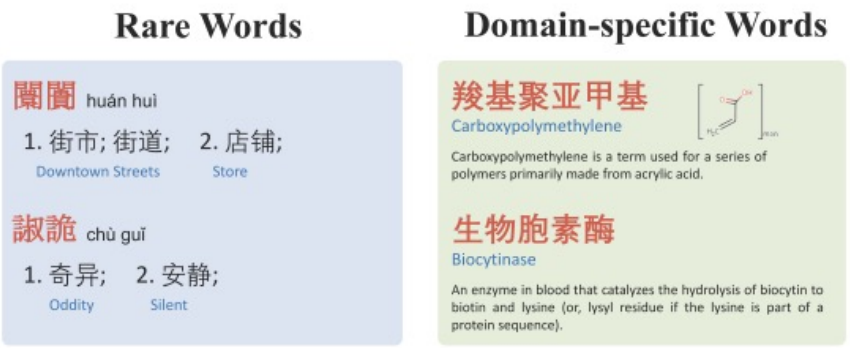


Figure 1: Examples of rare words and domain-specific words in dictionaries.

Figure 1: An example of multimodal tweets. In this tweet, "*Rocky*" is the name of the dog.

对任务的理解—经验

实验结果错误分析

大规模论文阅读--发现解决方案

Figure 1: An example of multimodal tweets.Without visual information, we can hardly predict the correct tag: #dog.

Hashtag Recommendation for Multimodal Microblog Using Co-Attention Network

CVPR, Visual Question Answering 任务论文

大规模论文阅读--任务转换

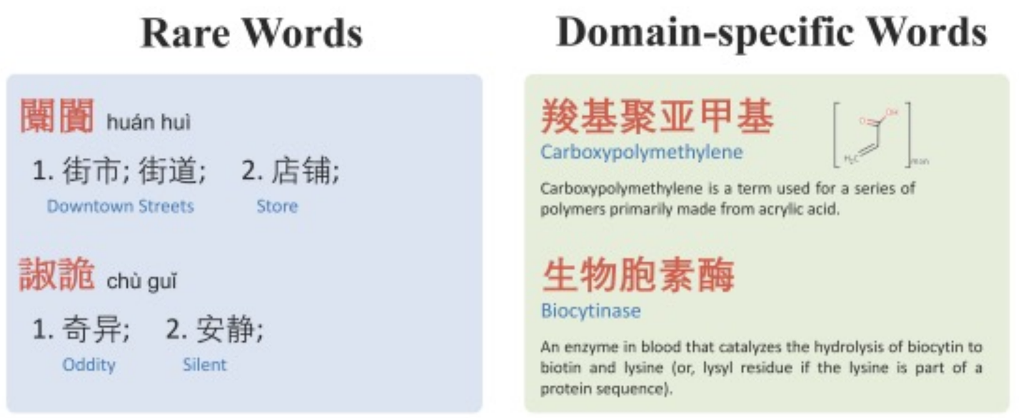


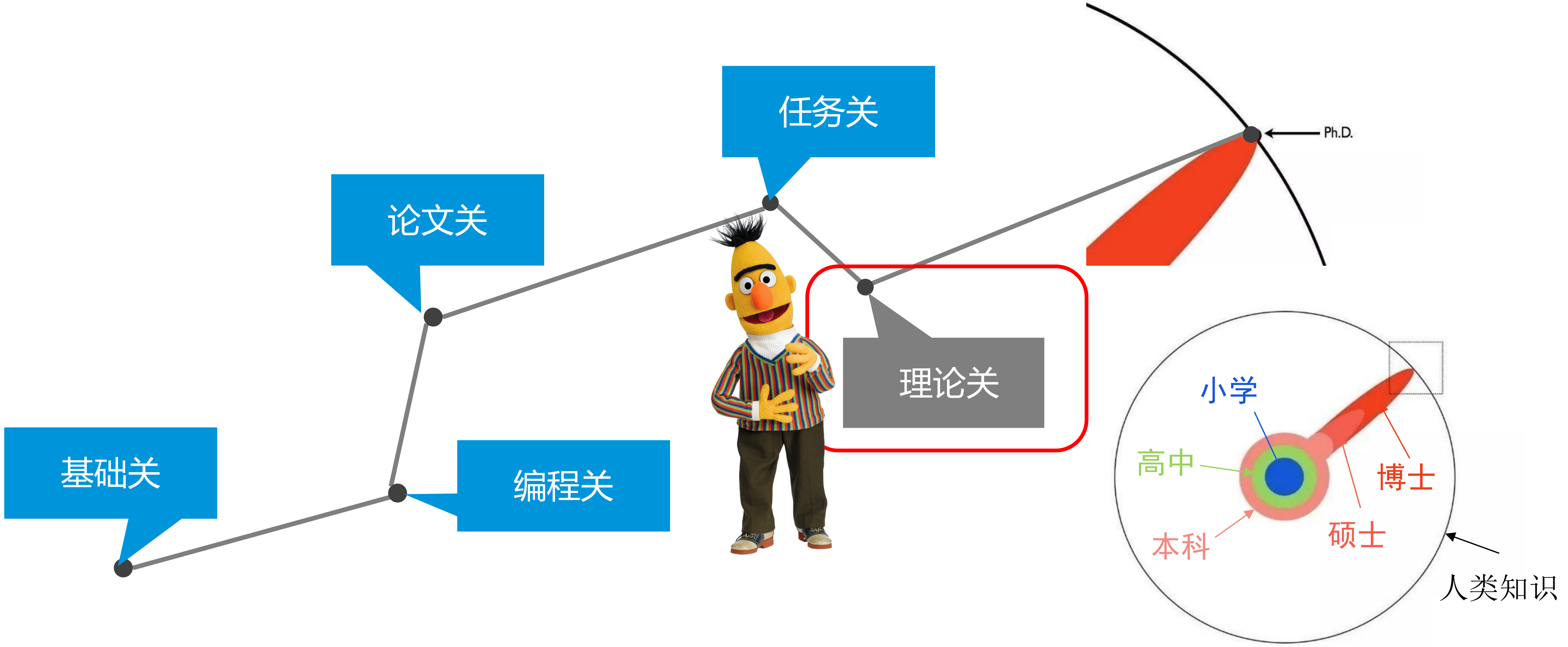
Figure 1: Examples of rare words and domain-specific words in dictionaries.

Figure 1: An example of multimodal tweets. In this tweet, "*Rocky*" is the name of the dog.

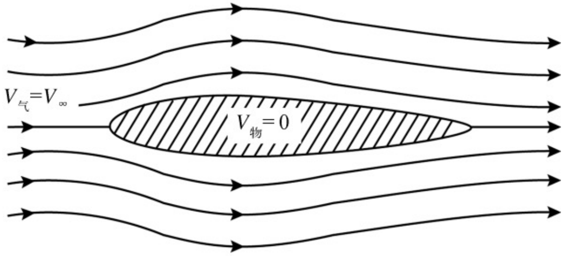
之前的分词算法在处理未登录词或者处理领域特有词时，会有问题

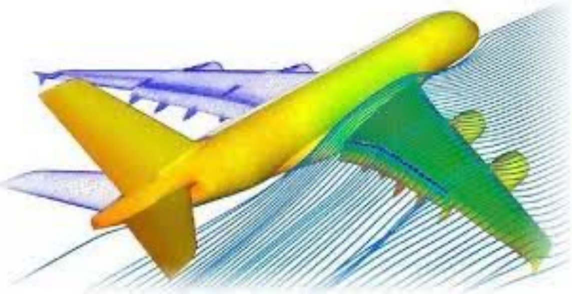
之前的NER算法不会关注图像信息，但是没有图像信息的话，一些Tweets不能很好分析

0 NLPer的打怪升级路线图



4 有了空气动力学才有了航空器的大发展





尝试

通用

理论

莱特兄弟飞行者一号 人类歷史上首次重于空气的航空器持续而且受控的动力飞行

第一代：基于规则的方法

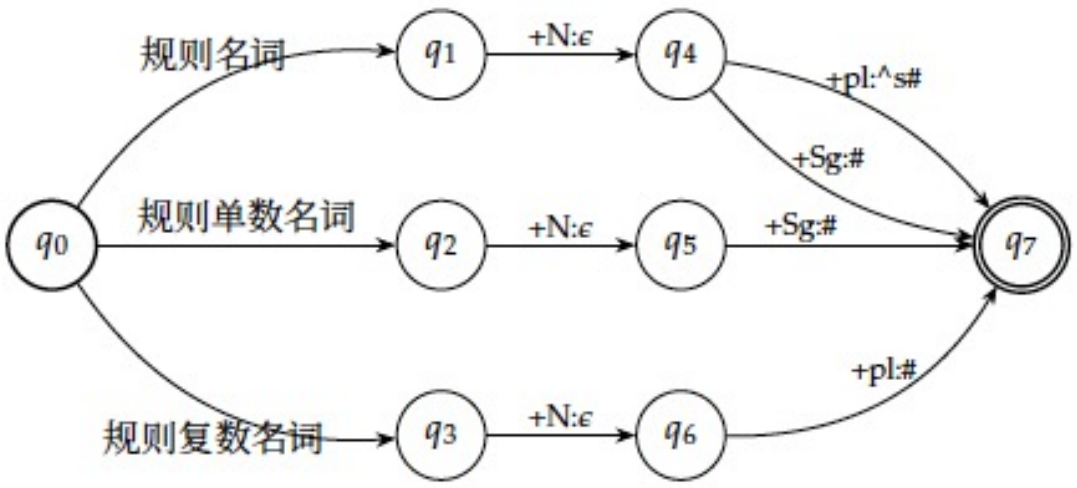
第二代：基于统计机器学习的方法

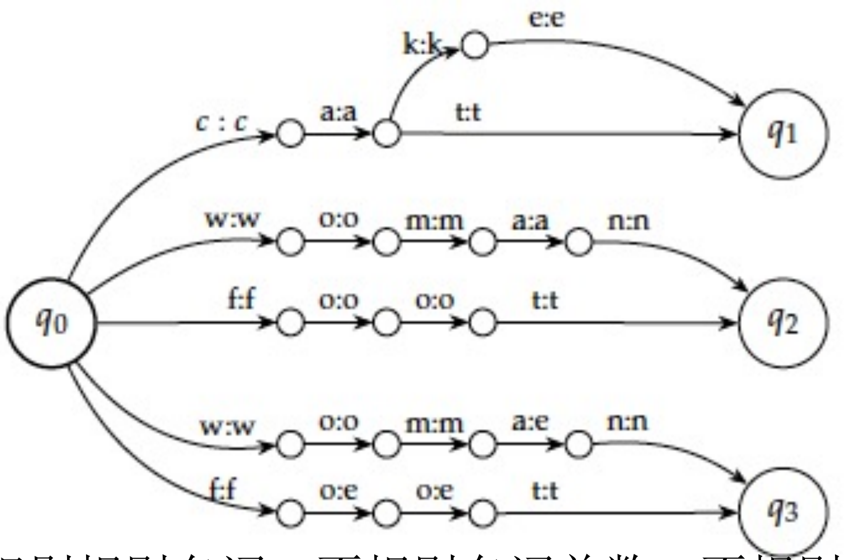
第三代：基于预训练的深度神经网络方法

第四代：基于超大规模预训练语言的 Prompt Learning？

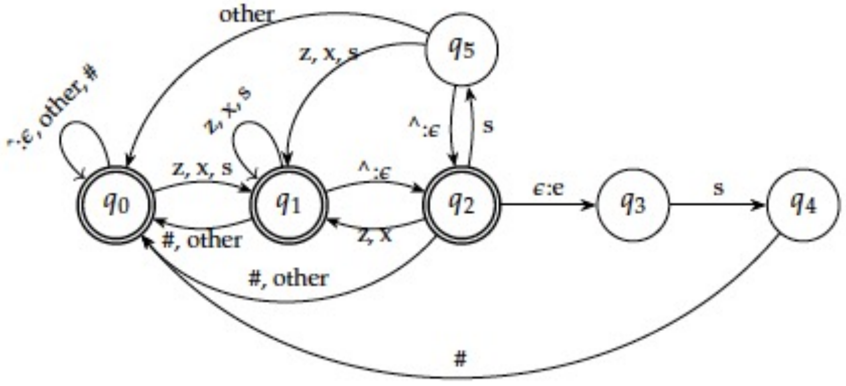
第一代：基于规则的方法

词形分析任务





识别规则名词、不规则名词单数、不规则名词复数



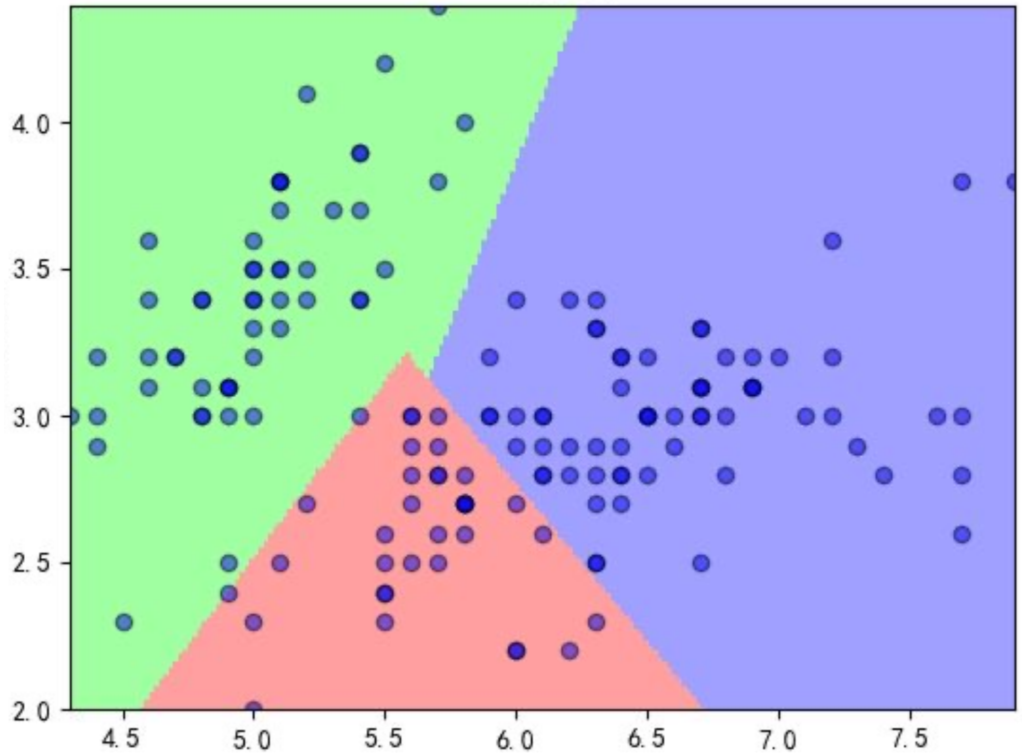
规则名词 “box”(以及其他以 “z，s，x” 结尾的规则名词)

第二代：基于统计机器学习的方法

中文分词任务

今晚	的	长安街	流光溢彩	。
BE	S	BIE	BIIE	S

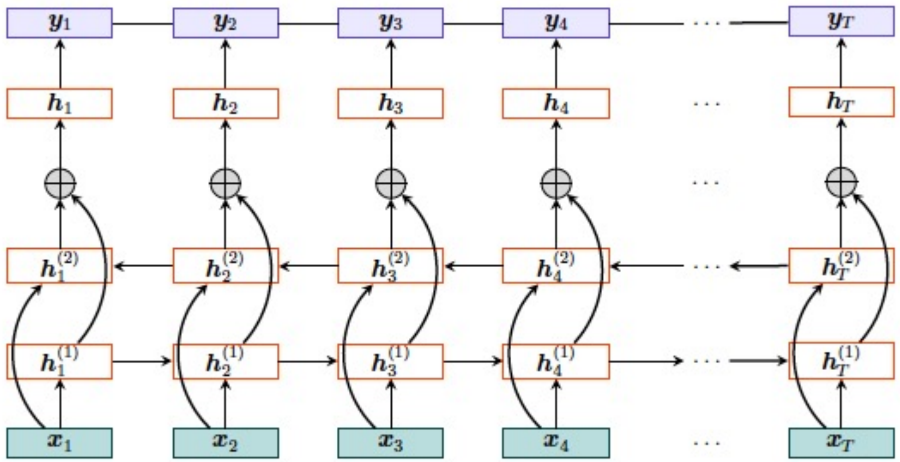
模板名	描述	例子
x_i	当前字	安
x_{i-1}	$i-1$ 位置的字	长
x_{i-2}	$i-2$ 位置的字	的
x_{i+1}	$i+1$ 位置的字	街
x_{i+2}	$i+2$ 位置的字	流
x_{i-2}, x_{i-1}	$i-2$ 开始的 bigram	长安
x_{i-1}, x_i	$i-1$ 开始的 bigram	的长
x_i, x_{i+1}	i 开始的 bigram	安街
x_{i+1}, x_{i+2}	$i+1$ 开始的 bigram	街流

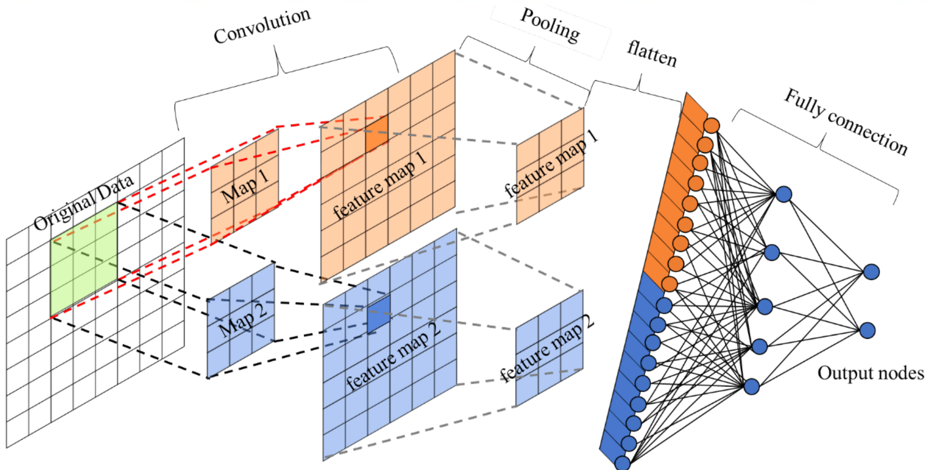


特征工程（Feature Engineering）：我们要进行“无聊“的特征模板定义环节

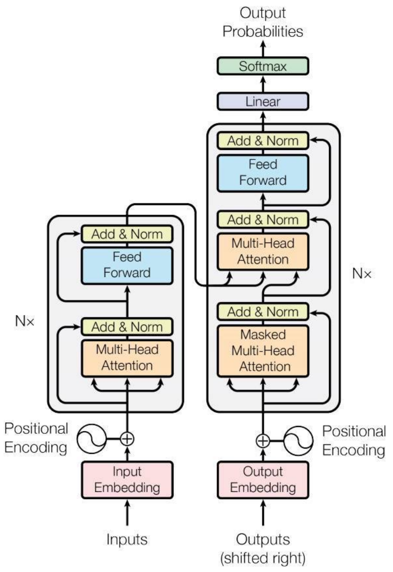
第二代：基于统计机器学习的方法

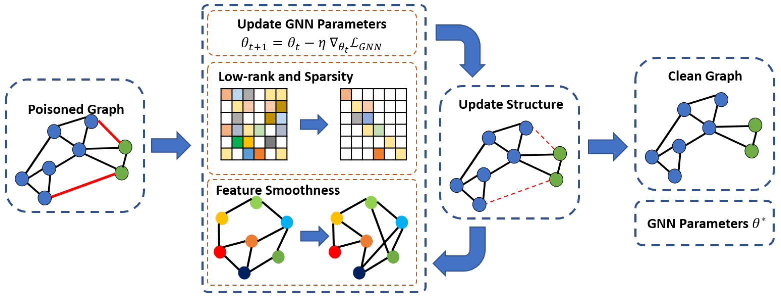
中文分词任务





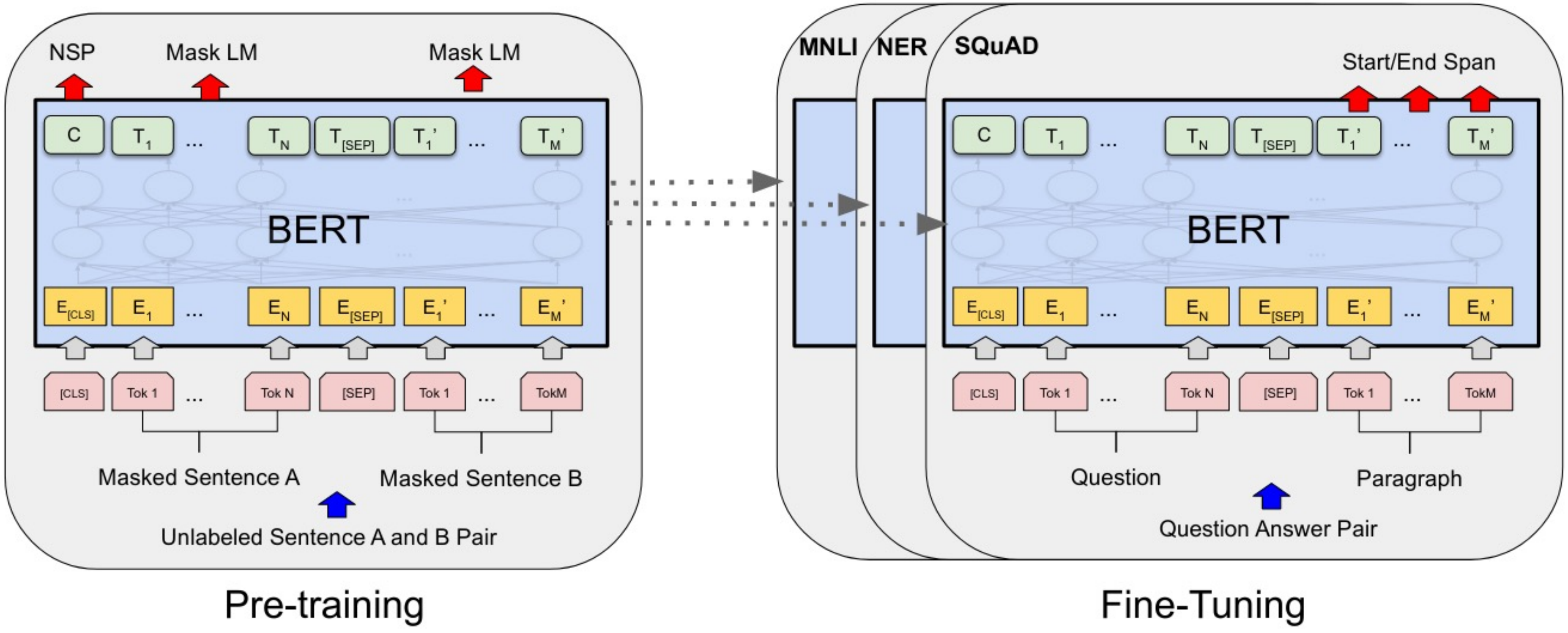
结构工程 （Architecture Engineering）：神经网络虽然解放手动配置特征模板所需要的人力，但是是以需要人工去设计合适网络结构。





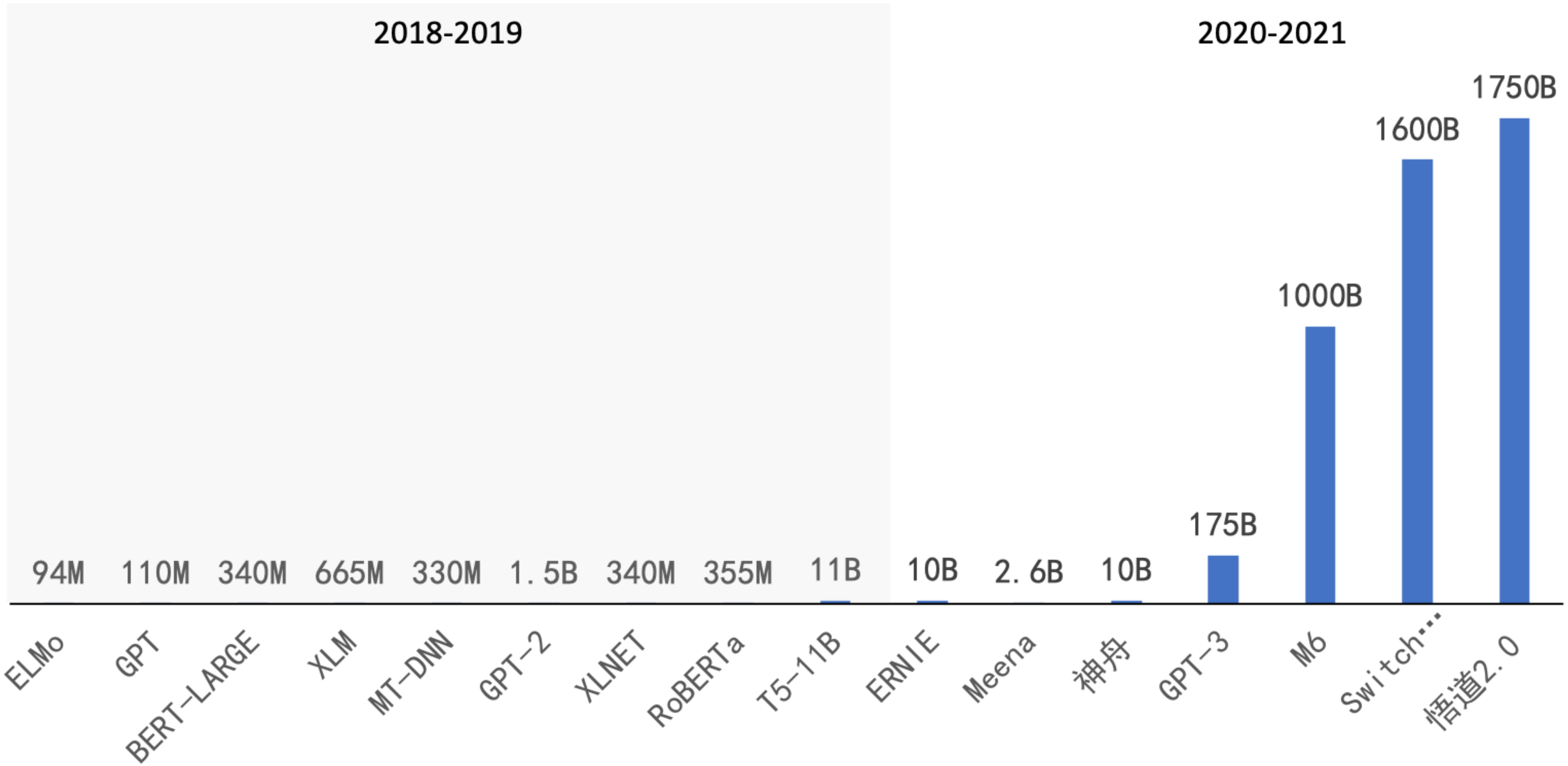
4 自然语言处理范式

第三代：基于预训练的深度神经网络方法

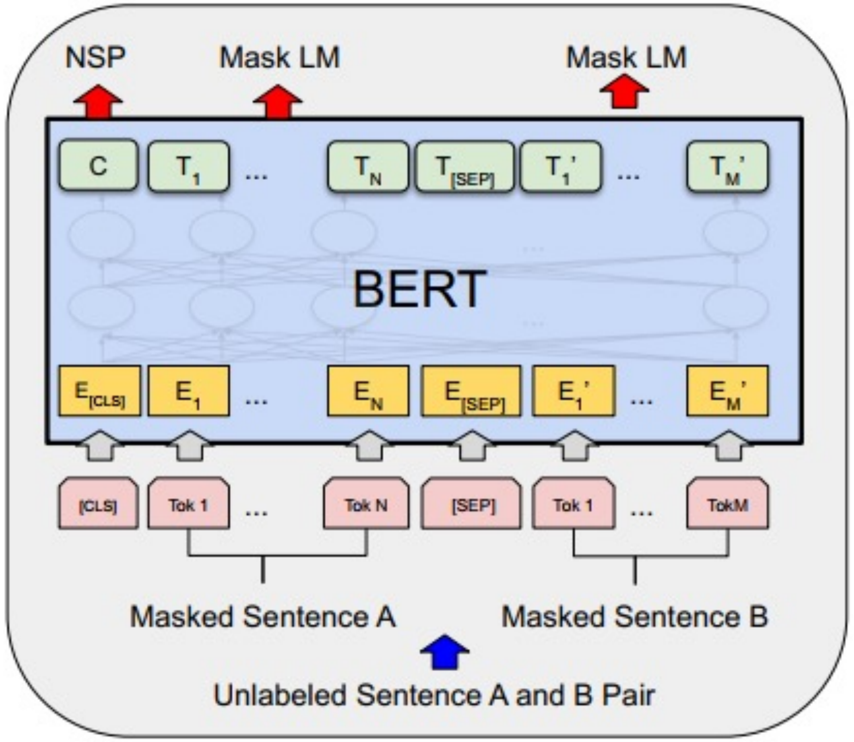


目标函数挖掘（Objective Engineering）： 这个过程研究者往往是通过引入额外的目标函数到预训练语言模型上，以便让其更适配下游任务

第三代：基于预训练的深度神经网络方法



第四代：基于超大规模预训练语言的 Prompt Learning？



Pre-training

情感倾向分析任务

输入：我喜欢这个电影

输出：“正面" 或者 "负面"

Prompt Learning转换为“完形填空"，

输入：我喜欢这个电影，*整体上来看，这是一个__ 的电影*

输出：“有趣的" 或者 "无聊的"

第一代：基于规则的方法　　语言学***

第二代：基于统计机器学习的方法　　语言学**，概率**，最优化***

第三代：基于预训练的深度神经网络方法　　语言学，概率*，最优化**

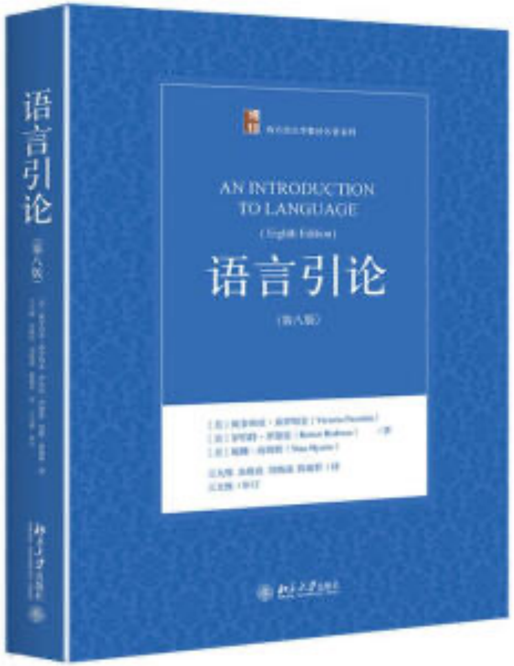
第四代：基于超大规模预训练语言的 Prompt Learning？　　语言学，概率*，最优化**

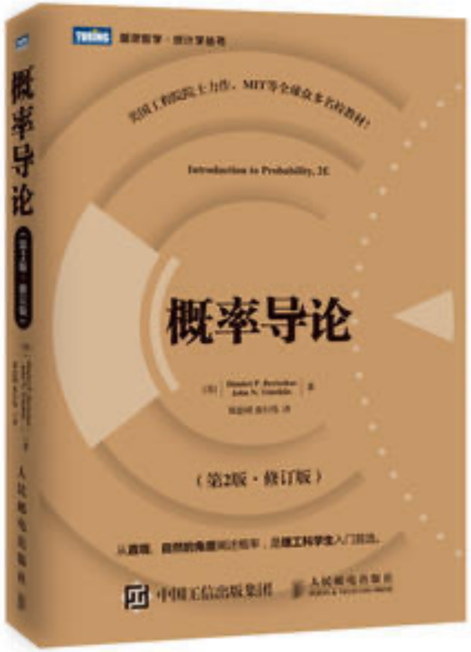
第N代：基于贝叶斯网络？

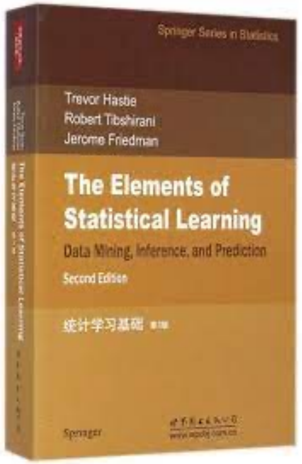
语言学，概率***，最优化**

第N+1代：基于因果推断？

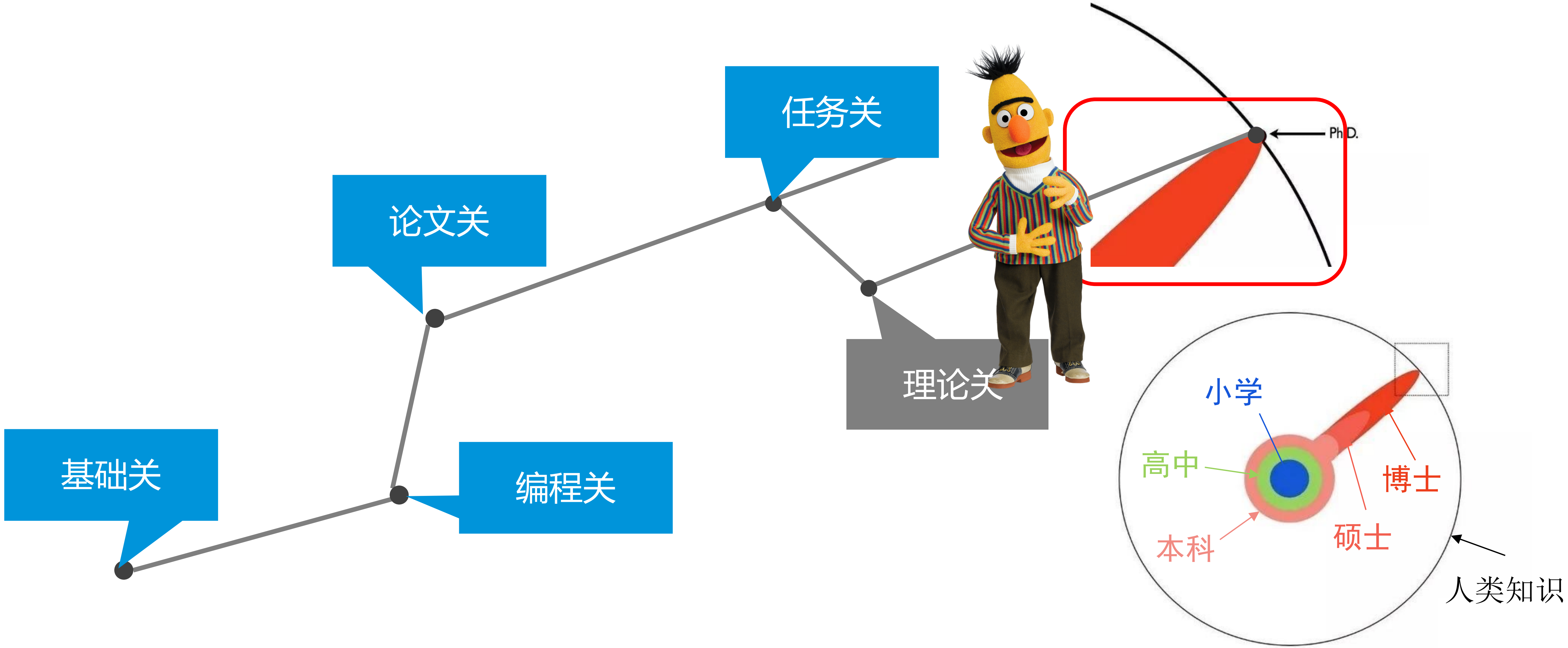
4 从事自然语言处理深层次研究需要掌握基础理论







0 NLPer的打怪升级路线图



5 如何写论文--- 论文文发表流程

确定方向	信息抽取
确定问题	如何提升命名实体识别的鲁棒性
确定思路	降低模型对实体名的依赖
确定方法	在损失函数上加入某种正则化项
实验验证	数据集、基线系统、评价指标
撰写论文	投稿ACL

通常大家认为的审稿人应该是这样审稿的：

审稿人一定是专家，无无所不知。打印出来，仔细研读揣摩数天，对于看不懂的地方反复推敲。即使你的英文文写得极其糟糕、即使你的文文章组织很混乱、即使你的表述很难看懂，审稿人花费了大量的时间后终于看懂了，他认为你的工作是有意义的，决定给你个border line或以上的分数。

实际上审稿人往往是这样审稿的：

他不一定是专家，一直忙于其他事，在deadline到来之前一天要完成n篇。审稿时他往往先看题目、摘要，扫一下introduction（知道你做什么），然后直接翻到最后找核心实验结果（做得好不好），然后基本确定录还是不录（也许只用5分钟！）。如果决定录，剩下就是写些赞美的话，指出些次要的小毛毛病。如果决定拒，下面的过程就是细看中间部分找理由拒了。

第一印象定录拒，5分钟内打动审稿人

5 如何写论文 – 微博上的佐证

胡云华MSRA + 加关注

最近有很多论文需要评审，跟同行聊天，得出一个有意思的结论：如果一篇论文在看完abstract和conclusion后还不能判断论文是否有价值的话，基本上这篇论文也就悲剧了。自己试了多次，屡试不爽。最极端的一篇我看了整整两天，全部搞懂作者在说什么后，仍然觉得应该拒掉，就跟只看5分钟得出的结论一致。

胡云华MSRA：回复@shirlywang1983:我说的是"小论文"，毕业论文之类的评审得少，不好说。好的论文需要准确提炼观点，让读者在尽量短的时间内明白你做了什么，你的贡献是什么。如果自己没想清楚，肯定写不清楚的。当然这个过程很不容易，没有深厚积累谁都做不到。 (12月5日 09:01)

kingdy9：说明第一印象很重要，也很准确。。有了第一印象后再找找文章中值得批判的地方就好了。。 //@朱小燕THU: 悲哀的是，已经感觉到了，但是为了写评语还是要看到底 (12月5日 09:38)

王伟DL：回复@胡云华MSRA:谢谢！我得修正我的观点，很同意"审论文时，abstract和conclusion写不好但内容好的情况少之又少。" (12月5日 14:22)

如何写论文—转变观念

以作者为核心整理工作

⇩

以读者为核心阐述工工作

信息的呈现符合读者的认知惯性

深入浅出，引人入胜，让读者快速找到想要的信息

尽量降低读者的理解难度

合理地综合使用信息元素：图>曲线>表>正文>公式

尽量提高读者阅读时的愉悦感

思想新颖、组织合理、逻辑严密
论证充分、文笔优美、排版美观

5 如何写论文—降低信息理解难度是关键

step	action	rule	stack	coverage
0				○○○○○○○
1	S	r_3	[The President will]	●●○○○○○
2	S	r_1	[The President will] [visit]	●●○○○○●
3	R_l		[The President will visit]	●●○○○○●
4	S	r_4	[The President will visit] [London in April]	●●●●●●●
5	R_r		[The President will visit London in April]	●●●●●●●

图

System	Setting	English–French	Chinese–English
GIZA++	Model 4 s2t	7.7	20.9
	Model 4 t2s	9.2	30.3
	Intersection	6.8	21.8
	Union	9.6	28.1
	Refined method	5.9	18.4
Cross-EM	HMM, joint	5.1	18.9
Vigne	Model 4 s2t	7.8	20.5
	+Model 4 t2s	5.6	18.3
	+link count	5.5	17.7
	+cross count	5.4	17.6
	+neighbor count	5.2	17.4
	+exact match	5.3	-
	+linked word count	5.2	17.3
	+bilingual dictionary	-	17.1
	+link co-occurrence count (GIZA++)	5.1	16.3
	+link co-occurrence count (Cross-EM)	4.0	15.7

表格

Shift-reduce parsing is efficient but suffers from parsing errors caused by syntactic ambiguity. Figure 3 shows two (partial) derivations for a dependency tree. Consider the item on the top, the algorithm can either apply a shift action to move a new item or apply a reduce left action to obtain a bigger structure. This is often referred to as **conflict** in the shift-reduce dependency parsing literature (Huang et al., 2009). In this work, the shift-reduce parser faces four types of conflicts:

正文

$$\frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_k} = \sum_{i=1}^{I}\sum_{\mathbf{y}\in\mathcal{Y}(\mathbf{x}^{(i)})} P(\mathbf{y}|\mathbf{x}^{(i)};\boldsymbol{\theta})\phi_k(\mathbf{x}^{(i)},\mathbf{y}) - \sum_{\mathbf{x}\in\mathcal{X}}\sum_{\mathbf{y}\in\mathcal{Y}(\mathbf{x})} P(\mathbf{x},\mathbf{y};\boldsymbol{\theta})\phi_k(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{I}\mathbb{E}_{\mathbf{y}|\mathbf{x}^{(i)};\boldsymbol{\theta}}[\phi_k(\mathbf{x}^{(i)},\mathbf{y})] - \mathbb{E}_{\mathbf{x},\mathbf{y};\boldsymbol{\theta}}[\phi_k(\mathbf{x},\mathbf{y})]$$

公式

Algorithm 1 A beam search algorithm for word alignment

procedure ALIGN($\mathbf{f}, \mathbf{e}$)
  $open \leftarrow \emptyset$ ▷ a list of active alignments
  $\mathcal{N} \leftarrow \emptyset$ ▷ n-best list
  $\mathbf{a} \leftarrow \emptyset$ ▷ begin with an empty alignment
  ADD($open, \mathbf{a}, \beta, b$) ▷ initialize the list
  while $open \neq \emptyset$ **do**
    $closed \leftarrow \emptyset$ ▷ a list of promising alignments
    for all $\mathbf{a} \in open$ **do**
      for all $l \in J \times I - \mathbf{a}$ **do** ▷ enumerate all possible new links
        $\mathbf{a}' \leftarrow \mathbf{a} \cup \{l\}$ ▷ produce a new alignment
        $g \leftarrow$ GAIN($\mathbf{f}, \mathbf{e}, \mathbf{a}, l$) ▷ compute the link gain
        if $g > 0$ **then** ▷ ensure that the score will increase
          ADD($closed, \mathbf{a}', \beta, b$) ▷ update promising alignments
        end if
        ADD($\mathcal{N}, \mathbf{a}', 0, n$) ▷ update n-best list
      end for
    end for
    $open \leftarrow closed$ ▷ update active alignments
  end while
  return $\mathcal{N}$ ▷ return n-best list
end procedure

算法

Proof of Theorem 1: Let $\bar{\alpha}^k$ be the weights before the k'th mistake is made. It follows that $\bar{\alpha}^1 = 0$. Suppose the k'th mistake is made at the i'th example. Take z to the output proposed at this example, $z = \arg\max_{y\in\mathbf{GEN}(x_i)} \Phi(x_i, y)\cdot\bar{\alpha}^k$. It follows from the algorithm updates that $\bar{\alpha}^{k+1} = \bar{\alpha}^k + \Phi(x_i, y_i) - \Phi(x_i, z)$. We take inner products of both sides with the vector $\mathbf{U}$:

$$\mathbf{U}\cdot\bar{\alpha}^{k+1} = \mathbf{U}\cdot\bar{\alpha}^k + \mathbf{U}\cdot\Phi(x_i, y_i) - \mathbf{U}\cdot\Phi(x_i, z) \geq \mathbf{U}\cdot\bar{\alpha}^k + \delta$$

where the inequality follows because of the property of $\mathbf{U}$ assumed in Eq. 3. Because $\bar{\alpha}^1 = 0$, and therefore $\mathbf{U}\cdot\bar{\alpha}^1 = 0$, it follows by induction on k that for all k, $\mathbf{U}\cdot\bar{\alpha}^{k+1} \geq k\delta$. Because $\mathbf{U}\cdot\bar{\alpha}^{k+1} \leq ||\mathbf{U}||\ ||\bar{\alpha}^{k+1}||$, it follows that $||\bar{\alpha}^{k+1}|| \geq k\delta$.

证明

如何看浩如烟海的文献？

- 根据标题过滤50%
- 根据摘要再过滤20%
- 根据介绍再过滤20%
- 剩下的10%再仔细看论文

黄铠

中科院云计算中心首席科学家
IEEE Life Fellow

5 如何写论文--标题的例子

Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data

John Lafferty[†*] LAFFERTY@CS.CMU.EDU
Andrew McCallum[*†] MCCALLUM@WHIZBANG.COM
Fernando Pereira[*‡] FPEREIRA@WHIZBANG.COM

[*]WhizBang! Labs–Research, 4616 Henry Street, Pittsburgh, PA 15213 USA
[†]School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213 USA
[‡]Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA 19104 USA

- 用一句话概括你所做的工作
- 考虑搜索引擎的影响，包含关键词

Cooooooooooooooollllllllllllll!!!!!!!!!!!!!!
Using Word Lengthening to Detect Sentiment in Microblogs

Samuel Brody
School of Communication
and Information
Rutgers University
sdbrody@gmail.com

Nicholas Diakopoulos
School of Communication
and Information
Rutgers University
diakop@rutgers.edu

- 可以适当地别出心裁

Who Will You "@"?

Yeyun Gong, Qi Zhang, Xuyang Sun, Xuanjing Huang
Shanghai Key Laboratory of Data Science
School of Computer Science, Fudan University
825 Zhangheng Road, Shanghai, P.R.China
{12110240006, qz,13210240106, xjhuang}@fudan.edu.cn

几句话概括你的工作

误区

- 力图把所有细节都说清楚
- 用很专业的术语来描述
- 出现数学符号

用语要简单，让外行能看懂

5 如何写论文--摘要的一个例子

Abstract

Distant supervision for relation extraction provides uniform bag labels for each sentence inside the bag, while accurate sentence labels are important for downstream applications that need the exact relation type. Directly using bag labels for sentence-level training will introduce much noise, thus severely degrading performance. In this work, we propose the use of negative training (NT), in which a model is trained using complementary labels regarding that "the instance does not belong to these complementary labels". Since the probability of selecting a true label as a complementary label is low, NT provides less noisy information. Furthermore, the model trained with NT is able to separate the noisy data from the training data. Based on NT, we propose a sentence-level framework, SENT, for distant relation extraction. SENT not only filters the noisy data to construct a cleaner dataset, but also performs a re-labeling process to transform the noisy data into useful training data, thus further benefiting the model's performance. Experimental results show the significant improvement of the proposed method over previous methods on sentence-level evaluation and de-noise effect.

研究的任务是什么

现有的方法有什么问题

我们的解决思路是什么

实现这个思路会遇到哪些难点

我们如何解决的

我们解决的还不错

Ma et al. Sentence-level Distant Relation Extraction via Negative Training, ACL 2021

5 如何写论文--摘要的一个例子

Abstract

Conventional *n*-best reranking techniques often suffer from the limited scope of the *n*-best list, which rules out many potentially good alternatives. We instead propose *forest reranking*, a method that reranks a packed forest of exponentially many parses. Since exact inference is intractable with non-local features, we present an approximate algorithm inspired by forest rescoring that makes discriminative training practical over the whole Treebank. Our final result, an F-score of 91.7, outperforms both 50-best and 100-best reranking baselines, and is better than any previously reported systems trained on the Treebank.

问题是什么

我们的解决思路是什么

我们如何解决的

我们解决的还不错

Liang Huang. Forest Reranking: Discriminative Parsing with Non-Local Features. In ACL 2008.

5 如何写论文—Introduction部分的写法

SENT: Sentence-level Distant Relation Extraction via Negative Training

Ruotian Ma[1], Tao Gui[2*], Linyang Li[1], Qi Zhang[1*], Xuanjing Huang[1] and Yaqian Zhou[1]
[1]School of Computer Science, Fudan University, Shanghai, China
[2]Institute of Modern Languages and Linguistics, Fudan University, Shanghai, China
{rtma19,tgui16,linyangli19,qz,xjhuang,yqzhou}@fudan.edu.cn

Abstract

Distant supervision for relation extraction provides uniform bag labels for each sentence inside the bag, while accurate sentence labels are important for downstream applications that need the exact relation type. Directly using bag labels for sentence-level training will introduce much noise, thus severely degrading performance. In this work, we propose the use of negative training (NT), in which a model is trained using complementary labels regarding that "the instance does not belong to these complementary labels". Since the probability of selecting a true label as a complementary label is low, NT provides less noisy information. Furthermore, the model trained with NT is able to separate the noisy data from the training data. Based on NT, we propose a sentence-level framework, SENT, for distant relation extraction. SENT not only filters the noisy data to construct a cleaner dataset, but also performs a relabeling process to transform the noisy data into useful training data, thus further benefiting the model's performance. Experimental results show the significant improvement of the proposed method over previous methods on sentence-level evaluation and de-noise effect.

1 Introduction

Relation extraction (RE), which aims to extrac
the relation between entity pairs from unstructure
text, is a fundamental task in natural language
processing. The extracted relation facts can benefi
various downstream applications, e.g., knowledge
graph completion (Bordes et al., 2013; Wang et al.,
2014), information extraction (Wu and Weld, 2010
and question answering (Yao and Van Durme
2014; Fader et al., 2014).

A significant challenge for relation extraction is the lack of large-scale labeled data. Thus, distant

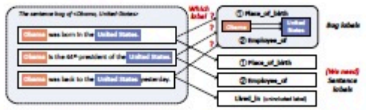


Figure 1: Two types of noise exist in bag-level labels: 1) Multi-label noise: the exact label ("place_of_birth" or "employee_of") for each sentence is unclear; 2) Wrong-label noise: the third sentence inside the bag actually expresses "live_in" which is not included in the bag labels.

supervision (Mintz et al., 2009) is proposed to gather training data through automatic alignment between a database and plain text. Such annotation paradigm results in an inevitable noise problem, which is alleviated by previous studies using multi-instance learning (MIL). In MIL, the training and testing processes are performed at the bag level, where a bag contains noisy sentences mentioning the same entity pair but possibly not describing the same relation. Studies using MIL can be broadly classified into two categories: 1) the soft de-noise methods that leverage soft weights to differentiate the influence of each sentence (Lin et al., 2016; Han et al., 2018c; Li et al., 2020; Hu et al., 2019a; Ye and Ling, 2019; Yuan et al., 2019a,b); 2) the hard de-noise methods that remove noisy sentences from the bag (Zeng et al., 2015; Qin et al., 2018; Han et al., 2018a; Shang, 2019).

However, these bag-level approaches fail to map each sentence inside bags with explicit sentence labels. This problem limits the application of RE in some downstream tasks that require sentence-level relation type, e.g., Yao and Van Durme (2014) and Xu et al. (2016) use sentence-level relation extraction to identify the relation between the answer and the entity in the question. Therefore, several studies (Jia et al. (2019); Feng et al. (2018)) have made efforts on sentence-level (or instance-level)

* Corresponding authors.

Relation extraction (RE), which aims to extract the relation between entity pairs from unstructured text, is a fundamental task in natural language processing. The extracted relation facts can benefit various downstream applications, e.g., knowledge graph completion (Bordes et al., 2013; Wang et al., 2014), information extraction (Wu and Weld, 2010) and question answering (Yao and Van Durme, 2014; Fader et al., 2014).

研究的任务是什么

5 如何写论文—Introduction部分的写法

SENT: Sentence-level Distant Relation Extraction via Negative Training

Ruotian Ma[1], Tao Gui[2*], Linyang Li[1], Qi Zhang[1*], Xuanjing Huang[1] and Yaqian Zhou[1]
[1]School of Computer Science, Fudan University, Shanghai, China
[2]Institute of Modern Languages and Linguistics, Fudan University, Shanghai, China
{rtma19,tgui16,linyangli19,qz,xjhuang,yqzhou}@fudan.edu.cn

Abstract

Distant supervision for relation extraction provides uniform bag labels for each sentence inside the bag, while accurate sentence labels are important for downstream applications that need the exact relation type. Directly using bag labels for sentence-level training will introduce much noise, thus severely degrading performance. In this work, we propose the use of negative training (NT), in which a model is trained using complementary labels regarding that "the instance does not belong to these complementary labels". Since the probability of selecting a true label as a complementary label is low, NT provides less noisy information. Furthermore, the model trained with NT is able to separate the noisy data from the training data. Based on NT, we propose a sentence-level framework, SENT, for distant relation extraction. SENT not only filters the noisy data to construct a cleaner dataset, but also performs a re-labeling process to transform the noisy data into useful training data, thus further benefiting the model's performance. Experimental results show the significant improvement of the proposed method over previous methods on sentence-level evaluation and de-noise effect.

1 Introduction

Relation extraction (RE), which aims to extract the relation between entity pairs from unstructured text, is a fundamental task in natural language processing. The extracted relation facts can benefit various downstream applications, e.g., knowledge graph completion (Bordes et al., 2013; Wang et al., 2014), information extraction (Wu and Weld, 2010) and question answering (Yao and Van Durme, 2014; Fader et al., 2014).

A significant challenge for relation extraction is the lack of large-scale labeled data. Thus, distant

* Corresponding authors.

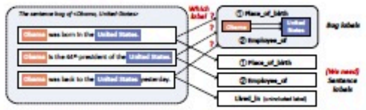


Figure 1: Two types of noise exist in bag-level labels: 1) Multi-label noise: the exact label ("place_of_birth" or "employee_of") for each sentence is unclear; 2) Wrong-label noise: the third sentence inside the bag actually expresses "live_in" which is not included in the bag labels.

supervision (Mintz et al., 2009) is proposed to gather training data through automatic alignment between a database and plain text. Such annotation paradigm results in an inevitable noise problem, which is alleviated by previous studies using multi-instance learning (MIL). In MIL, the training and testing processes are performed at the bag level, where a bag contains noisy sentences mentioning the same entity pair but possibly not describing the same relation. Studies using MIL can be broadly classified into two categories: 1) the soft de-noise methods that leverage soft weights to differentiate the influence of each sentence (Lin et al., 2016; Han et al., 2018c; Li et al., 2020; Hu et al., 2019a; Ye and Ling, 2019; Yuan et al., 2019a,b); 2) the hard de-noise methods that remove noisy sentences from the bag (Zeng et al., 2015; Qin et al., 2018; Han et al., 2018a; Shang, 2019).

However, these bag-level approaches fail to map each sentence inside bags with explicit sentence labels. This problem limits the application of RE in some downstream tasks that require sentence-level relation type, e.g., Yao and Van Durme (2014) and Xu et al. (2016) use sentence-level relation extraction to identify the relation between the answer and the entity in the question. Therefore, several studies (Jia et al. (2019); Feng et al. (2018)) have made efforts on sentence-level (or instance-level)

A significant challenge for relation extraction is the lack of large-scale labeled data. Thus, distant supervision (Mintz et al., 2009) is proposed to gather training data through automatic alignment between a database and plain text. Such annotation paradigm results in an inevitable noise problem, which is alleviated by previous studies using multi-instance learning (MIL). In MIL, the training and testing processes are performed at the bag level, where a bag contains noisy sentences mentioning the same entity pair but possibly not describing the same relation. Studies using MIL can be broadly classified into two categories: 1) the soft de-noise methods that leverage soft weights to differentiate the influence of each sentence (Lin et al., 2016; Han et al., 2018c; Li et al., 2020; Hu et al., 2019a; Ye and Ling, 2019; Yuan et al., 2019a,b); 2) the hard de-noise methods that remove noisy sentences from the bag (Zeng et al., 2015; Qin et al., 2018; Han et al., 2018a; Shang, 2019).

现有的方法哪些

SENT: Sentence-level Distant Relation Extraction via Negative Training

Ruotian Ma[1], Tao Gui[2*], Linyang Li[1], Qi Zhang[1*], Xuanjing Huang[1] and Yaqian Zhou[1]
[1]School of Computer Science, Fudan University, Shanghai, China
[2]Institute of Modern Languages and Linguistics, Fudan University, Shanghai, China
{rtma19,tgui16,linyangli19,qz,xjhuang,yqzhou}@fudan.edu.cn

Abstract

Distant supervision for relation extraction provides uniform bag labels for each sentence inside the bag, while accurate sentence labels are important for downstream applications that need the exact relation type. Directly using bag labels for sentence-level training will introduce much noise, thus severely degrading performance. In this work, we propose the use of negative training (NT), in which a model is trained using complementary labels regarding that "the instance does not belong to these complementary labels". Since the probability of selecting a true label as a complementary label is low, NT provides less noisy information. Furthermore, the model trained with NT is able to separate the noisy data from the training data. Based on NT, we propose a sentence-level framework, SENT, for distant relation extraction. SENT not only filters the noisy data to construct a cleaner dataset, but also performs a re-labeling process to transform the noisy data into useful training data, thus further benefiting the model's performance. Experimental results show the significant improvement of the proposed method over previous methods on sentence-level evaluation and de-noise effect.

1 Introduction

Relation extraction (RE), which aims to extract the relation between entity pairs from unstructured text, is a fundamental task in natural language processing. The extracted relation facts can benefit various downstream applications, e.g., knowledge graph completion (Bordes et al., 2013; Wang et al., 2014), information extraction (Wu and Weld, 2010) and question answering (Yao and Van Durme, 2014; Fader et al., 2014).

A significant challenge for relation extraction is the lack of large-scale labeled data. Thus, distant

* Corresponding authors.

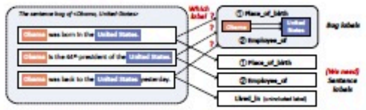


Figure 1: Two types of noise exist in bag-level labels: 1) Multi-label noise: the exact label ("place_of_birth" or "employee_of") for each sentence is unclear; 2) Wrong-label noise: the third sentence inside the bag actually expresses "live_in" which is not included in the bag labels.

supervision (Mintz et al., 2009) is proposed to gather training data through automatic alignment between a database and plain text. Such annotation paradigm results in an inevitable noise problem, which is alleviated by previous studies using multi-instance learning (MIL). In MIL, the training and testing processes are performed at the bag level, where a bag contains noisy sentences mentioning the same entity pair but possibly not describing the same relation. Studies using MIL can be broadly classified into two categories: 1) the soft de-noise methods that leverage soft weights to differentiate the influence of each sentence (Lin et al., 2016; Han et al., 2018c; Li et al., 2020; Hu et al., 2019a; Ye and Ling, 2019; Yuan et al., 2019a,b); 2) the hard de-noise methods that remove noisy sentences from the bag (Zeng et al., 2015; Qin et al., 2018; Han et al., 2018a; Shang, 2019).

However, these bag-level approaches fail to map each sentence inside bags with explicit sentence labels. This problem limits the application of RE in some downstream tasks that require sentence-level relation type, e.g., Yao and Van Durme (2014) and Xu et al. (2016) use sentence-level relation extraction to identify the relation between the answer and the entity in the question. Therefore, several studies (Jia et al. (2019); Feng et al. (2018)) have made efforts on sentence-level (or instance-level)

However, these bag-level approaches fail to map each sentence inside bags with explicit sentence labels. This problem limits the application of RE in some downstream tasks that require sentence-level relation type, e.g., Yao and Van Durme (2014) and Xu et al. (2016) use sentence-level relation extraction to identify the relation between the answer and the entity in the question. Therefore, several studies (Jia et al. (2019); Feng et al. (2018)) have made efforts on sentence-level (or instance-level)

现有的方法有什么问题

5 如何写论文—Introduction部分的写法

distant RE, empirically verifying the deficiency of bag-level methods on sentence-level evaluation. However, the instance selection approaches of these methods depend on rewards(Feng et al., 2018) or frequent patterns(Jia et al., 2019) determined by bag-level labels, which contain much noise. For one thing, one bag might be assigned to multiple bag labels, leading to difficulties in one-to-one mapping between sentences and labels. As shown in Fig.1, we have no access to the exact relation between "place_of_birth" and "employee_of" for the sentence "Obama was born in the United States.". For another, the sentences inside a bag might not express the bag relations. In Fig.1, the sentence "Obama was back to the United States yesterday" actually express the relation "live_in", which is not included in the bag labels.

In this work, we propose the use of negative training (NT) (Kim et al., 2019) for distant RE. Different from positive training (PT), NT trains a model by selecting the complementary labels of the given label, regarding that "the input sentence does not belong to this complementary label". Since the probability of selecting a true label as a complementary label is low, NT decreases the risk of providing noisy information and prevents the model from overfitting the noisy data. Moreover, the model trained with NT is able to separate the noisy data from the training data (a histogram in Fig.3 shows the separated data distribution during NT). Based on NT, we propose SENT, a sentence-level framework for distant RE. During SENT training, the noisy instances are not only filtered with a noise-filtering strategy, but also transformed into useful training data with a re-labeling method. We further design an iterative training algorithm to take full advantage of these data-refining processes, which significantly boost performance. Our codes are publicly available at *Github*[1].

To summarize the contribution of this work:

- We propose the use of negative training for sentence-level distant RE, which greatly protects the model from noisy information.
- We present a sentence-level framework, SENT, which includes a noise-filtering and a re-labeling strategy for re-fining distant data.
- The proposed method achieves significant improvement over previous methods in terms of both RE performance and de-noise effect.

[1] https://github.com/rtmaww/SENT

2 Related Work

2.1 Distant Supervision for RE

Supervised relation extraction (RE) has been constrained by the lack of large-scale labeled data. Therefore, distant supervision (DS) is introduced by Mintz et al. (2009), which employs existing knowledge bases (KBs) as source of supervision instead of annotated text. Riedel et al. (2010) relaxes the DS assumption to the express-at-least-once assumption. As a result, multi-instance learning is introduced (Riedel et al. (2010); Hoffmann et al. (2011); Surdeanu et al. (2012)) for this task, where the training and evaluating process are performed in **bag-level**, with potential noisy sentences existing in each bag. Most following studies in distant RE adopt this paradigm, aiming to decrease the impact of noisy sentences in each bag. These studies include the attention-based methods to attend to useful information (Lin et al. (2016); Han et al. (2018c); Li et al. (2020); Hu et al. (2019a); Ye and Ling (2019); Yuan et al. (2019a); Zhu et al. (2019); Yuan et al. (2019b); Wu et al. (2017)), the selection strategies such as RL or adversarial training to remove noisy sentences from the bag (Zeng et al. (2015); Shang (2019); Qin et al. (2018); Han et al. (2018a)) and the incorporation with extra information such as KGs, multi-lingual corpora or other information (Ji et al. (2017); Lei et al. (2018); Vashishth et al. (2018); Han et al. (2018b); Zhang et al. (2019); Qu et al. (2019); Verga et al. (2016); Lin et al. (2017); Wang et al. (2018); Deng and Sun (2019); Beltagy et al. (2019)). Other approaches include soft-label strategy for denoising (Liu et al. (2017)), leveraging pre-trained LM (Alt et al. (2019)), pattern-based method (Zheng et al. (2019)), structured learning method (Bai and Ritter (2019)) and so forth (Luo et al. (2017); Chen et al. (2019)).

In this work, we focus on **sentence-level** relation extraction. Several previous studies also perform Distant RE on sentence-level. Feng et al. (2018) proposes a reinforcement learning framework for sentence selecting, where the reward is given by the classification scores on bag labels. Jia et al. (2019) builds an initial training set and further select confident instances based on selected patterns. The difference between the proposed work and previous works is that we do not rely on bag-level labels for sentence selecting. Furthermore, we leverage NT to dynamically separate the noisy data from

In this work, we propose the use of negative training (NT) (Kim et al., 2019) for distant RE. Different from positive training (PT), NT trains a model by selecting the complementary labels of the given label, regarding that "the input sentence does not belong to this complementary label". Since the probability of selecting a true label as a complementary label is low, NT decreases the risk of providing noisy information and prevents the model from overfitting the noisy data. Moreover, the model trained with NT is able to separate the noisy data from the training data (a histogram in Fig.3 shows the separated data distribution during NT). Based on NT, we propose SENT, a sentence-level framework for distant RE. During SENT training, the noisy instances are not only filtered with a noise-filtering strategy, but also transformed into useful training data with a re-labeling method. We further design an iterative training algorithm to take full advantage of these data-refining processes, which significantly boost performance. Our codes are publicly available at *Github*[1].

我们的解决思路是什么

我们如何解决的

5 如何写论文—Introduction部分的写法

distant RE, empirically verifying the deficiency of bag-level methods on sentence-level evaluation. However, the instance selection approaches of these methods depend on rewards(Feng et al., 2018) or frequent patterns(Jia et al., 2019) determined by bag-level labels, which contain much noise. For one thing, one bag might be assigned to multiple bag labels, leading to difficulties in one-to-one mapping between sentences and labels. As shown in Fig.1, we have no access to the exact relation between "place_of_birth" and "employee_of" for the sentence "Obama was born in the United States.". For another, the sentences inside a bag might not express the bag relations. In Fig.1, the sentence "Obama was back to the United States yesterday" actually express the relation "live_in", which is not included in the bag labels.

In this work, we propose the use of negative training (NT) (Kim et al., 2019) for distant RE. Different from positive training (PT), NT trains a model by selecting the complementary labels of the given label, regarding that "the input sentence does not belong to this complementary label". Since the probability of selecting a true label as a complementary label is low, NT decreases the risk of providing noisy information and prevents the model from overfitting the noisy data. Moreover, the model trained with NT is able to separate the noisy data from the training data (a histogram in Fig.3 shows the separated data distribution during NT). Based on NT, we propose SENT, a sentence-level framework for distant RE. During SENT training, the noisy instances are not only filtered with a noise-filtering strategy, but also transformed into useful training data with a re-labeling method. We further design an iterative training algorithm to take full advantage of these data-refining processes, which significantly boost performance. Our codes are publicly available at Github[1].

To summarize the contribution of this work:

- We propose the use of negative training for sentence-level distant RE, which greatly protects the model from noisy information.
- We present a sentence-level framework, SENT, which includes a noise-filtering and a re-labeling strategy for re-fining distant data.
- The proposed method achieves significant improvement over previous methods in terms of both RE performance and de-noise effect.

[1]https://github.com/rtmaww/SENT

2 Related Work

2.1 Distant Supervision for RE

Supervised relation extraction (RE) has been constrained by the lack of large-scale labeled data. Therefore, distant supervision (DS) is introduced by Mintz et al. (2009), which employs existing knowledge bases (KBs) as source of supervision instead of annotated text. Riedel et al. (2010) relaxes the DS assumption to the express-at-least-once assumption. As a result, multi-instance learning is introduced (Riedel et al. (2010); Hoffmann et al. (2011); Surdeanu et al. (2012)) for this task, where the training and evaluating process are performed in **bag-level**, with potential noisy sentences existing in each bag. Most following studies in distant RE adopt this paradigm, aiming to decrease the impact of noisy sentences in each bag. These studies include the attention-based methods to attend to useful information (Lin et al. (2016); Han et al. (2018c); Li et al. (2020); Hu et al. (2019a); Ye and Ling (2019); Yuan et al. (2019a); Zhu et al. (2019); Yuan et al. (2019b); Wu et al. (2017)), the selection strategies such as RL or adversarial training to remove noisy sentences from the bag (Zeng et al. (2015); Shang (2019); Qin et al. (2018); Han et al. (2018a)) and the incorporation with extra information such as KGs, multi-lingual corpora or other information (Ji et al. (2017); Lei et al. (2018); Vashishth et al. (2018); Han et al. (2018b); Zhang et al. (2019); Qu et al. (2019); Verga et al. (2016); Lin et al. (2017); Wang et al. (2018); Deng and Sun (2019); Beltagy et al. (2019)). Other approaches include soft-label strategy for denoising (Liu et al. (2017)), leveraging pre-trained LM (Alt et al. (2019)), pattern-based method (Zheng et al. (2019)), structured learning method (Bai and Ritter (2019)) and so forth (Luo et al. (2017); Chen et al. (2019)).

In this work, we focus on **sentence-level** relation extraction. Several previous studies also perform Distant RE on sentence-level. Feng et al. (2018) proposes a reinforcement learning framework for sentence selecting, where the reward is given by the classification scores on bag labels. Jia et al. (2019) builds an initial training set and further select confident instances based on selected patterns. The difference between the proposed work and previous works is that we do not rely on bag-level labels for sentence selecting. Furthermore, we leverage NT to dynamically separate the noisy data from

To summarize the contribution of this work:

- We propose the use of negative training for sentence-level distant RE, which greatly protects the model from noisy information.
- We present a sentence-level framework, SENT, which includes a noise-filtering and a re-labeling strategy for re-fining distant data.
- The proposed method achieves significant improvement over previous methods in terms of both RE performance and de-noise effect.

我们的核心贡献是什么

SENT: Sentence-level Distant Relation Extraction via Negative Training

Ruotian Ma[1], Tao Gui[2*], Linyang Li[1], Qi Zhang[1*], Xuanjing Huang[1] and Yaqian Zhou[1]
[1]School of Computer Science, Fudan University, Shanghai, China
[2]Institute of Modern Languages and Linguistics, Fudan University, Shanghai, China
{rtma19,tgui16,linyangli19,qz,xjhuang,yqzhou}@fudan.edu.cn

Abstract

Distant supervision for relation extraction provides uniform bag labels for each sentence inside the bag, while accurate sentence labels are important for downstream applications that need the exact relation type. Directly using bag labels for sentence-level training will introduce much noise, thus severely degrading performance. In this work, we propose the use of negative training (NT), in which a model is trained using complementary labels regarding that "the instance does not belong to these complementary labels". Since the probability of selecting a true label as a complementary label is low, NT provides less noisy information. Furthermore, the model trained with NT is able to separate the noisy data from the training data. Based on NT, we propose a sentence-level framework, SENT, for distant relation extraction. SENT not only filters the noisy data to construct a cleaner dataset, but also performs a re-labeling process to transform the noisy data into useful training data, thus further benefiting the model's performance. Experimental results show the significant improvement of the proposed method over previous methods on sentence-level evaluation and de-noise effect.

1 Introduction

Relation extraction (RE), which aims to extract the relation between entity pairs from unstructured text, is a fundamental task in natural language processing. The extracted relation facts can benefit various downstream applications, e.g., knowledge graph completion (Bordes et al., 2013; Wang et al., 2014), information extraction (Wu and Weld, 2010) and question answering (Yao and Van Durme, 2014; Fader et al., 2014).

A significant challenge for relation extraction is the lack of large-scale labeled data. Thus, distant

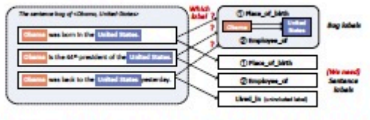


Figure 1: Two types of noise exist in bag-level labels: 1) Multi-label noise: the exact label ("place_of_birth" or "employee_of") for each sentence is unclear; 2) Wrong-label noise: the third sentence inside the bag actually expresses "live_in" which is not included in the bag labels.

supervision (Mintz et al., 2009) is proposed to gather training data through automatic alignment between a database and plain text. Such annotation paradigm results in an inevitable noise problem, which is alleviated by previous studies using multi-instance learning (MIL). In MIL, the training and testing processes are performed at the bag level, where a bag contains noisy sentences mentioning the same entity pair but possibly not describing the same relation. Studies using MIL can be broadly classified into two categories: 1) the soft de-noise methods that leverage soft weights to differentiate the influence of each sentence (Lin et al., 2016; Han et al., 2018c; Li et al., 2020; Hu et al., 2019a; Ye and Ling, 2019; Yuan et al., 2019a,b); 2) the hard de-noise methods that remove noisy sentences from the bag (Zeng et al., 2015; Qin et al., 2018; Han et al., 2018a; Shang, 2019).

However, these bag-level approaches fail to map each sentence inside bags with explicit sentence labels. This problem limits the application of RE in some downstream tasks that require sentence-level relation type, e.g., Yao and Van Durme (2014) and Xu et al. (2016) use sentence-level relation extraction to identify the relation between the answer and the entity in the question. Therefore, several studies (Jia et al. (2019); Feng et al. (2018)) have made efforts on sentence-level (or instance-level)

* Corresponding authors.

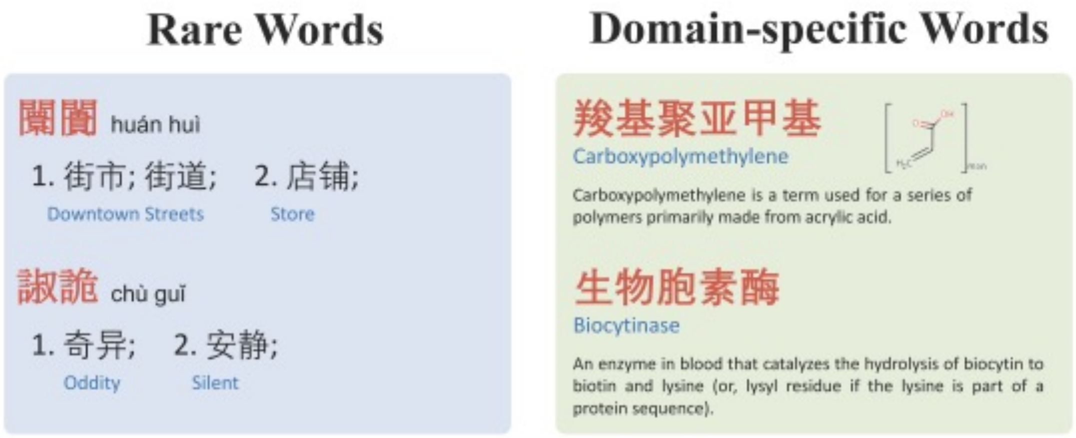


Figure 1: Examples of rare words and domain-specific words in dictionaries.

现在非常流行在第一页右上部分加入一个图表说明本文的**Motivation**

5 如何写论文—段落的写法

- 每个段落有个论断性的中心句
- 其余部分都是支撑句，围绕中心心句展开论证
  - 前人工作
  - 具体数据
- 支撑句之间可分类组织
- 段尾可以加上衔接句

The need to segment and label sequences arises in many different problems in several scientific fields. Hidden Markov models (HMMs) and stochastic grammars are well understood and widely used probabilistic models for such problems. In computational biology, HMMs and stochastic grammars have been successfully used to align biological sequences, find sequences homologous to a known evolutionary family, and analyze RNA secondary structure (Durbin et al., 1998). In computational linguistics and computer science, HMMs and stochastic grammars have been applied to a wide variety of problems in text and speech processing, including topic segmentation, part-of-speech (POS) tagging, information extraction, and syntactic disambiguation (Manning & Schütze, 1999).

John Lafferty, Andrew McCallum, and Fernando Pereira. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In ICML 2003.

不要一上来就描述你的工作，可以先介绍背景知识（往往就是baseline）

- 有利于降低初学者或其他领域学者的理解难度
- 有利于对introduction中的论文做更详细的解释
- 有利于对比baseline和你的方法

5 如何写论文--Running Example是利器

- 英语不好说不清楚？用例子！
- 全篇统一使用一个running example，用来阐释你的方法（甚至是baseline）
- 围绕着running example，展开描述你的工作
- 审稿人能从running example中更舒服地了解你的工作，读正文会花掉他/她更多时间
- 看完running example，审稿人便能知道核心思想

Figure 2: Example of feature vector construction. The character with the red shadow is the character x_i. The character segments with rounded rectangle are the words in the dictionary D.

- 公认的标准数据和state-of-the-art系统
- 实验先主后辅
    - 主实验（测试集）：证明显著超过baseline
    - 辅助实验（开发集）：参数的影响
- 不辞辛劳，做到极致！！！

5 如何写论文--Caption包含充分的信息

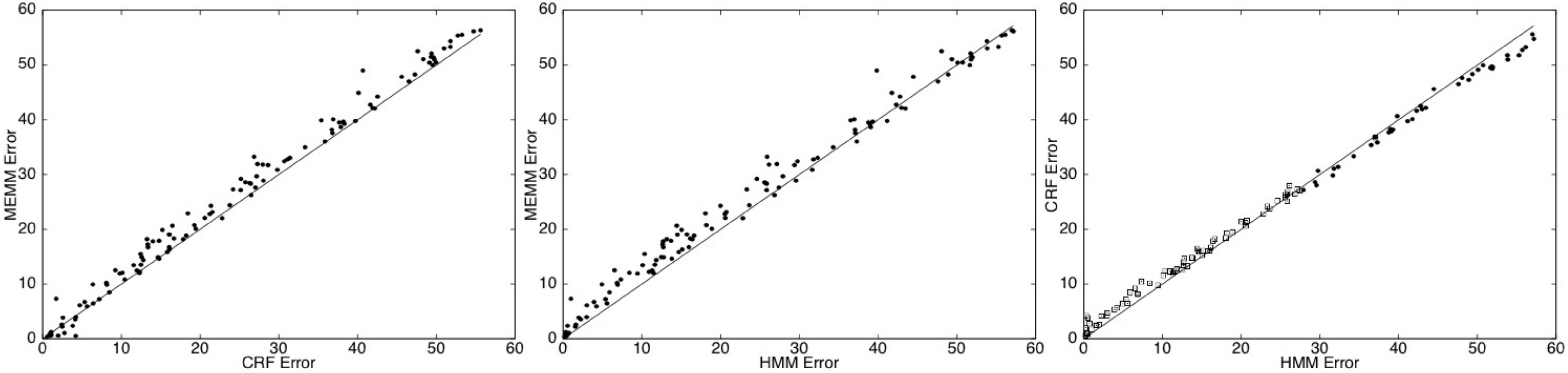


Figure 3. Plots of 2×2 error rates for HMMs, CRFs, and MEMMs on randomly generated synthetic data sets, as described in Section 5.2. As the data becomes "more second order," the error rates of the test models increase. As shown in the left plot, the CRF typically significantly outperforms the MEMM. The center plot shows that the HMM outperforms the MEMM. In the right plot, each open square represents a data set with $\alpha < \frac{1}{2}$, and a solid circle indicates a data set with $\alpha \geq \frac{1}{2}$. The plot shows that when the data is mostly second order ($\alpha \geq \frac{1}{2}$), the discriminatively trained CRF typically outperforms the HMM. These experiments are not designed to demonstrate the advantages of the additional representational power of CRFs and MEMMs relative to HMMs.

最好能直接看懂图，不用再去看正文

如何写相关工作

附录的写作技巧

写作常见问题

引用的写法

提高英语写作的窍门

CCIR 2017

学术论文写作技巧

刘洋

2017年7月，上海

1

更喜岷山千里雪，三军过后尽开颜。

谢谢！