# 当 NLP 邂逅 Social Media

## 我的小目标与大坚持

汇报人：桂韬

导师：黄萱菁、张奇教授

复旦大学自然语言处理实验室

想好一个 Idea

能否代码实现

资源够不够用

Idea 是否有效

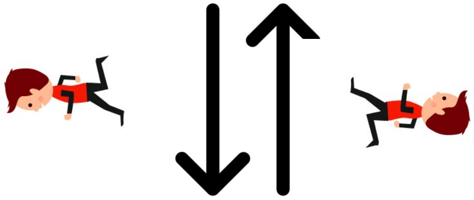
是否拼命三郎

论文写得好不好

Reviewer 是否认可

# 发论文的几道关卡

# 闯关失败



# 推倒重来

2018/6 EMNLP论文被接受

**2018/3 IJCAI论文被被拒**

**2018/2 ACL论文被被拒**

2017/11 AAAI论文被被拒

2017/6 EMNLP论文被接受

2017/3 ACL论文被被拒

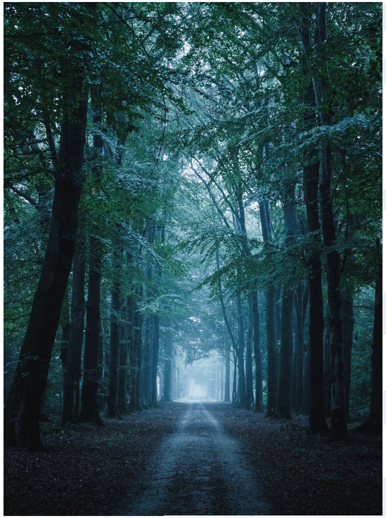
关键是坚持
坚持的动力是有自己的目标

- 啃一些硬骨头
- 把问题想得极端
- 打破惯性思维

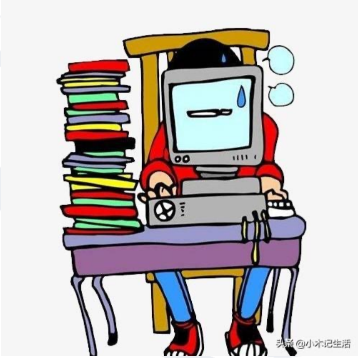
# 目标 1：啃一些硬骨头

开始一件事容易，坚持一件事困难

阳光大道



幽林秘境

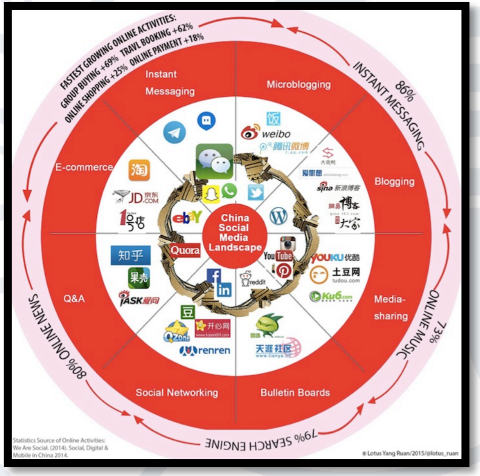



如何 **走进** NLP 大门

# 社交媒体

自发传播

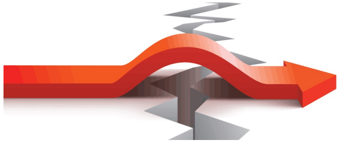

“社会化”属性




表现形式多样

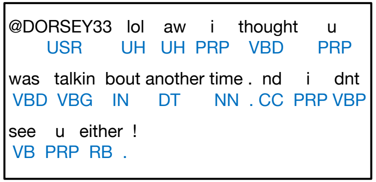
# 非规范性

明年他要 C 位出道
这是神马规矩
I 服了 U！
皮一下，很开心



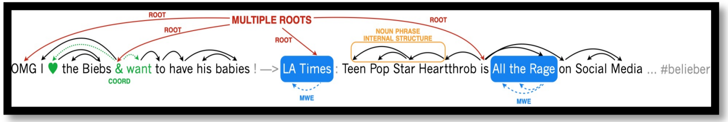

热点追踪

# Tweet POS Tagging




标注非常少


旧词新意、再造新词

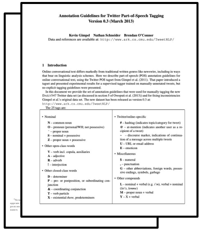



不遵循通常的语法、语用

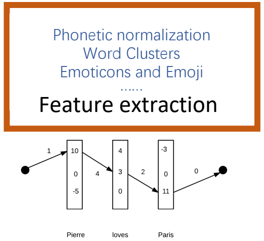
MEMM Tagger
+Twitter orthography
+Frequently-capitalized tokens
+Traditional tag dictionary
+ Distributional similarity
+ Phonetic normalization
+ Unsupervised Word Clusters
+ Emoticons and Emoji
+ Lexical Features

| Methods | RIT-Test |
|---|---|
| Stanford-WSJ (Toutanova et al., 2003) | 73.37% |
| Stanford-MIX | 83.14% |
| T-POS (Ritter et al., 2011) | 84.55% |
| GATE Tagger (Derczynski et al., 2013) | 88.69% |
| ARK Tagger (Owoputi et al., 2013) | 90.40% |
| bi-LSTM (word level) | 75.91% |

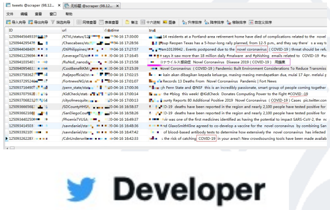
讨论可行方案



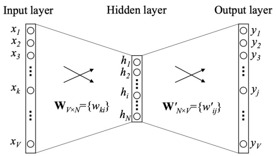

标注规范学习




机器学习方法复现

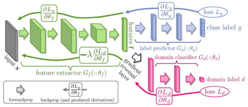



Twitter 数据爬取



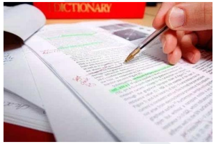

训练词向量


深度学习框架




对抗理论

开始训练




论文撰写


论文发表

每一篇论文都是汗水的结晶

图片来自https://kuaibao.qq.com/s/20190724A0N2U200?refer=spider

**Wall Street Journal Section of Penn Treebank**

As their varied strategies suggest , there is more than one way to respond to a disaster.

Treebank-3 (LDC1999T42) /07/WSJ_0799.POS

RT @jamstik : Lol :) there is more than one way to start living a greener life.

1

我们需要部分迁移

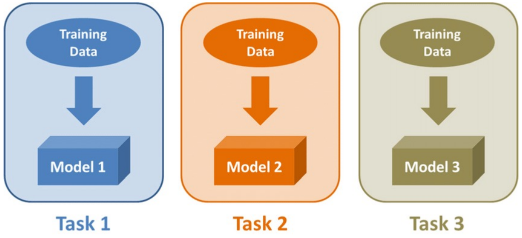

HyperNet　　开始训练　　论文撰写　　论文发表

有了前面的基础，更重要的是想出好的 idea

图片来自https://kuaibao.qq.com/s/20190724A0N2U200?refer=spider

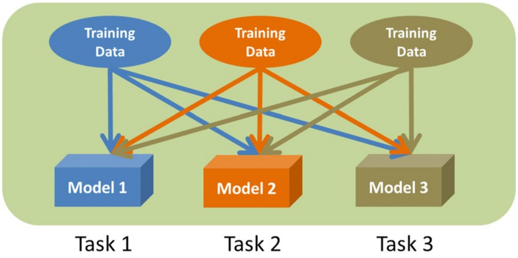
# 目标 2：把问题想到极端

真理往往掌握在少数人手中，坚持 idea 的闪光点很重要




(*a*)

# Single task learning




(*b*)

# Multi-task learning

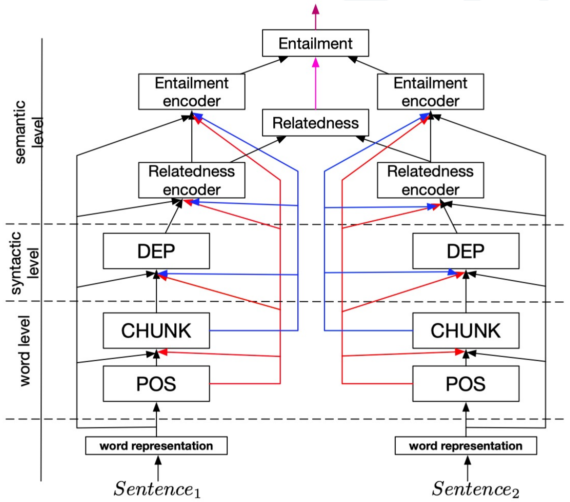



Figure 1: Overview of the joint many-task model predicting different linguistic outputs at successively deeper layers.

A Joint Many-Task Model: Growing a Neural Network for Multiple NLP Tasks, EMNLP 2017

Two difficulties:

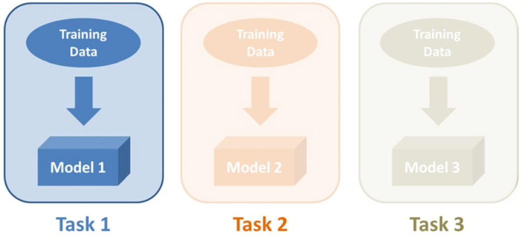
1. a sufficient number of related tasks
2. a sufficient number of related tasks

| | CCG | CHU | COM | FNT | POS | HYP | KEY | MWE | SEM | STR |
|---|---|---|---|---|---|---|---|---|---|---|
| CCG | | 1.4 | 0.45 | 0.58 | 1.8 | 0.24 | 0.3 | 0.45 | 1.4 | 0.84 |
| CHU | -0.052 | | -0.15 | -0.12 | -0.45 | -0.5 | -0.22 | -0.27 | -0.099 | -0.32 |
| COM | -5 | 1.3 | | 1.3 | -1.4 | -2.4 | -4.8 | 0.82 | -3 | -0.63 |
| FNT | -5.8 | -1 | -6.1 | | -9.4 | -5.7 | -3.6 | -9.4 | -3 | -0.68 |
| POS | 4.9 | 2.9 | 1.9 | 0.9 | | -0.85 | -0.26 | 1.3 | 3.4 | 2.9 |
| HYP | 12 | 4 | -11 | 9.2 | 22 | | 1.5 | -7.7 | 23 | 8.1 |
| KEY | 5.7 | 3.2 | -1 | -0.43 | -1.3 | -2.6 | | -4.7 | 0.59 | 0.69 |
| MWE | 18 | 20 | 7.4 | 5.5 | 1.6 | -3.8 | -5.8 | | 16 | 8.6 |
| SEM | -5 | -0.76 | -1.2 | -0.81 | -0.85 | -1.3 | -0.83 | -1.1 | | -1.7 |
| STR | -1.7 | 1.5 | -0.26 | -0.72 | 0.037 | -1.5 | -1.4 | -1.6 | 1.7 | |

Figure 1: Relative gains and losses (in percent) over main task micro-averaged $F_1$ when incorporating auxiliary tasks (columns) compared to single-task models for the main tasks (rows).

Identifying beneficial task relations for multi-task learning in deep neural networks, EACL 2017.

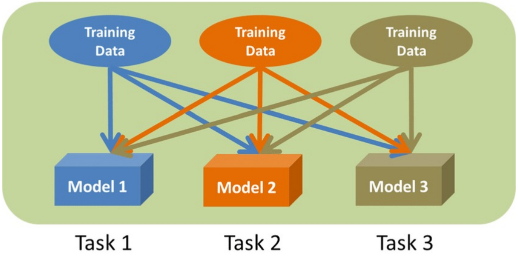
# 极端情况




(*a*)

Single task learning

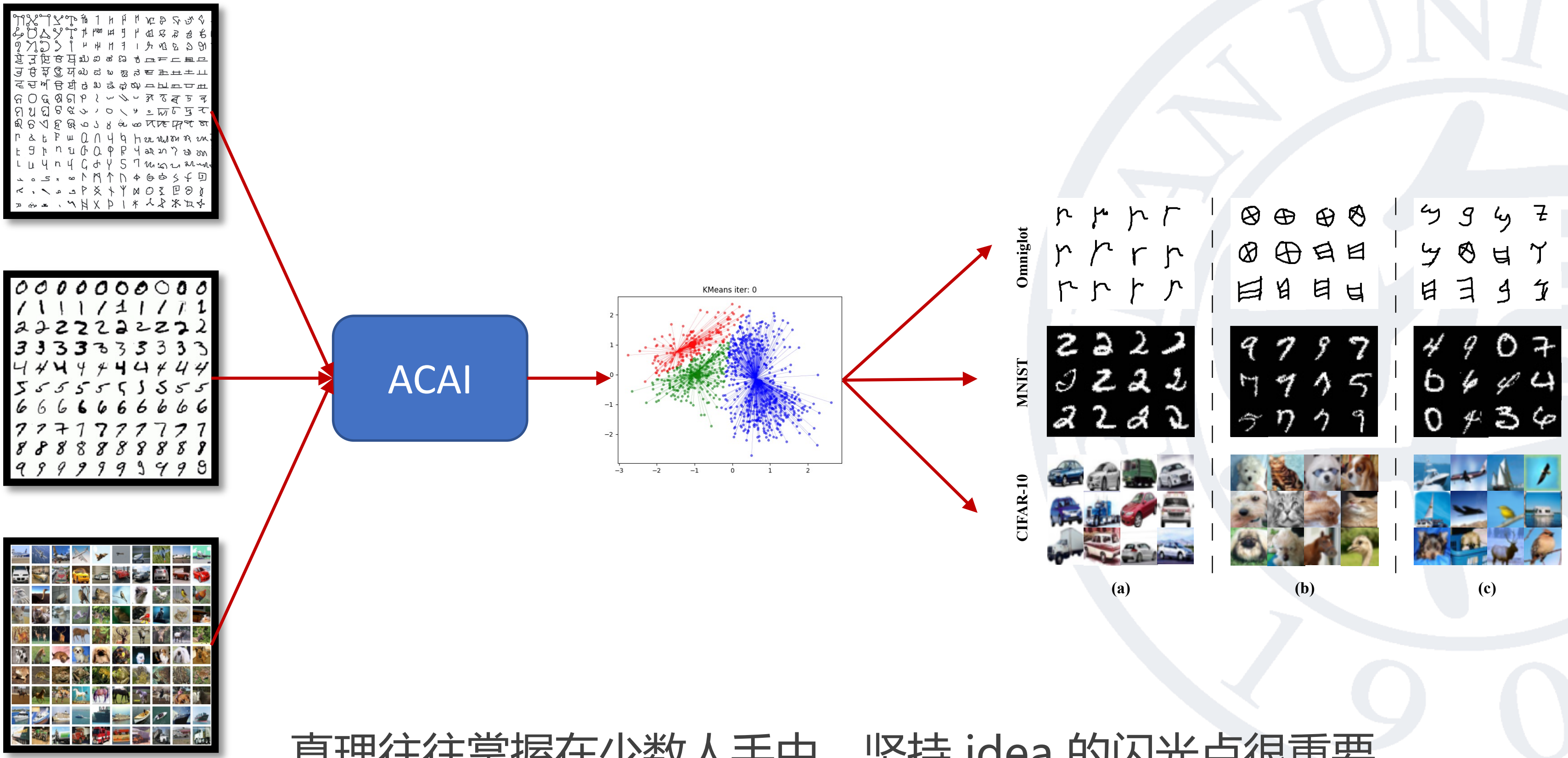
Only **one** task exists




(*b*)

Multi-task learning

How to construct
**Multiple Tasks for Augmentation**




真理往往掌握在少数人手中，坚持 idea 的闪光点很重要

**Algorithm 1** **Meta-MTL with $K$-means Augmentation**

Run embedding learning algorithm $\mathcal{E}$ on $D_{aux}$ and produce embeddings $\{\mathbf{z}_i\}$ from observations $\{\mathbf{x}_i\}$.
Run $k$-means on $\{\mathbf{z}_i\}$ $T$ times (with random scaling or random selection on dimensions) to generate a set of partitions $\{\mathcal{P}_t = \{C^l\}_{l=1}^{L_t}\}_{t=1}^{T}$, which correspond to a set of auxiliary tasks $\{\mathcal{T}_t\}_{t=1}^{T}$.
**for** episode = 1, M **do**
    Sample batch of tasks $\mathcal{T} \sim \{\mathcal{T}_t\}_{t=0}^{T}$.
    **for all** $\mathcal{T}$ **do**
        Sample $K$ datapoints $D_{\mathcal{T}} = \{\mathbf{x}_j, \mathbf{y}_j\}$.
        Evaluate $\nabla_{\theta_{\mathcal{F}}}$ and $\nabla_{\theta_{\mathcal{D}_t}}$ using $D_{\mathcal{T}}$ based on Equation 1.
        Applying gradient decent to update the parameters of task-specific decoders $\theta_{\mathcal{D}_{\mathcal{T}}}$.
        Compute updated parameters $\theta^*_{\mathcal{F}}$ with gradient descent based on Equation 5.
        Sample datapoints $D_0 = \{\mathbf{x}_j, \mathbf{y}_j\}$ from $\mathcal{T}_0$ for the meta-update.
    **end for**
    Update the parameters of shared layers $\theta_{\mathcal{F}}$ based on Equation 6.
**end for**

Big improvement in all aspects

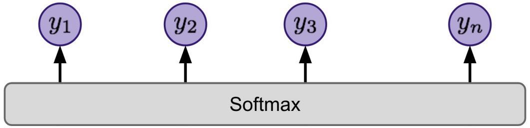
- Few-shot learning
- Semi-supervised learning
- Incorporate with data augmentation
- High-resolution images

Even the result increased by 10% accuracy

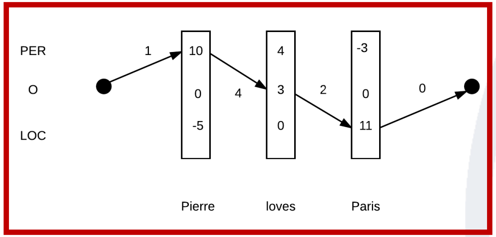
# 目标 3：打破惯性思维

坚持打破舒适圈

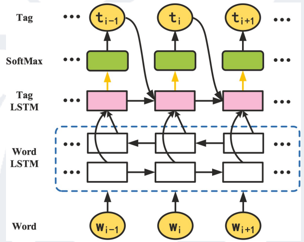
# As for Label Decoding Layer



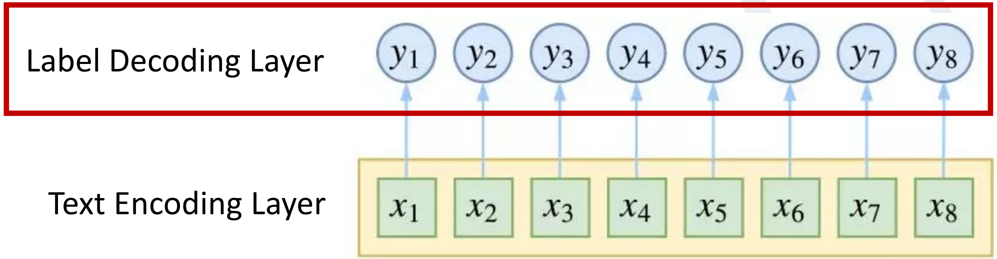

Softmax




CRF




Seq2seq



# As for Label Decoding Layer

| Decoding Methods | Strength | Weakness |
|---|---|---|
| Softmax | parallel decoding | No label dependency |
| CRF | Local label dependency | Viterbi  decoding |
| Seq2seq | Long-term label dependency | Sequence decoding |

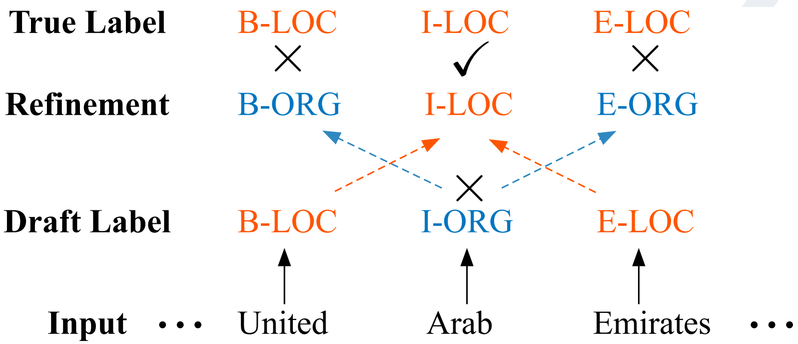
**Comparison of different label decoding methods**

# As for Label Decoding Layer

| Decoding Methods | Strength | Weakness |
| --- | --- | --- |
| Softmax | **parallel decoding** | No label dependency |
| CRF | Local label dependency | Viterbi decoding |
| Seq2seq | **Long-term label dependency** | Sequence decoding |

**What do we want?**

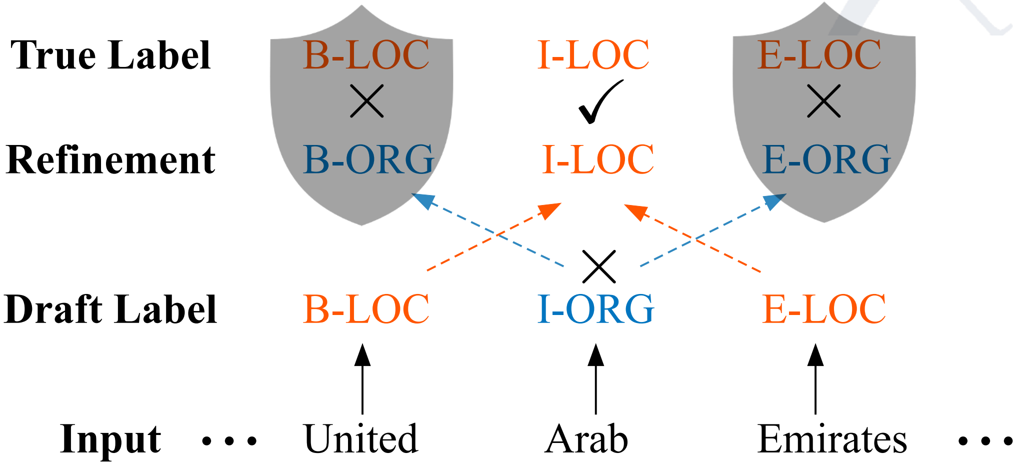
# Uncertainty-Aware Label Refinement for Sequence Labeling




| Refinement | #Tokens |
|---|---|
| ✔ → ✘ | **39** |
| ✘ → ✔ | 54 |

Table 1: Results of LAN with uncertainty estimation evaluated on CoNLL2003 test dataset. ✔ refers to the correct prediction, and ✘ refers to the wrong prediction.

# Uncertainty-Aware Label Refinement for Sequence Labeling

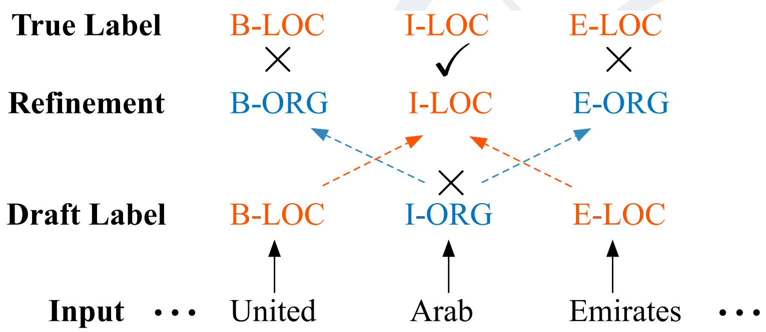


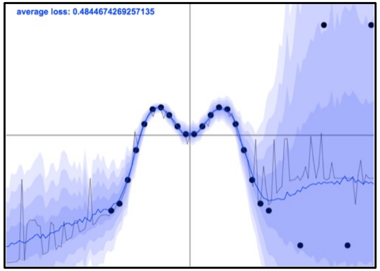

**Can we fine an indicator?**

Tao Gui, Jiacheng Ye, Qi Zhang, Zhengyan Li, zichu fei, Yeyun Gong and Xuanjing Huang
Uncertainty-Aware Label Refinement for Sequence Labeling, EMNLP 2020.

# Uncertainty-Aware Label Refinement for Sequence Labeling

## A Long Search




Bayesian NNs for
Uncertainty Estimation

| Draft | Uncertainty | Refinement | #Tokens |
|---|---|---|---|
| ✔ | 0.018 | ✔ → ✘ | **39** |
| ✘ | **0.524** | ✘ → ✔ | 54 |

Table 1: Results of LAN with uncertainty estimation evaluated on CoNLL2003 test dataset. ✔ refers to the correct prediction, and ✘ refers to the wrong prediction.

**Glove + DocL-NER 92.13**

| Models | $F_1$ |
|---|---|
| BERT-base [Devlin *et al.*, 2019] | 91.82* |
| **BERT-base + DocL-NER** | **92.92** |
| ELMo [Peters *et al.*, 2018] | 92.64* |
| **ELMo + DocL-NER** | **93.05** |

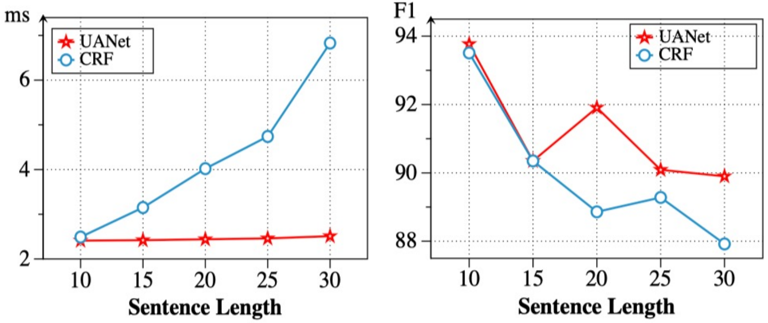
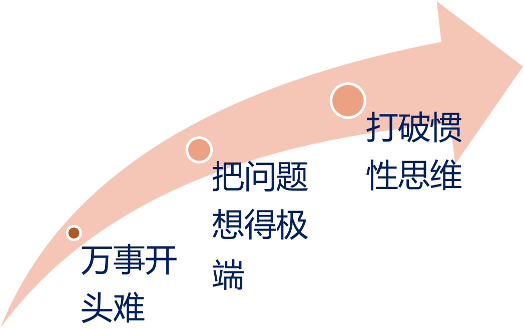
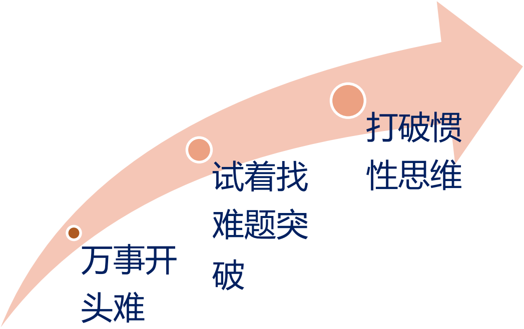



Figure 3: Speed and F1 against sentence length.




你觉得很累吗




如果你觉得很累

那是因为**你在走上坡路**