Words of experience

on scholar for beginners

——以NLP为例

费豪

2020/6/4

Content

- 写论文
  - 如何进入本领域
  - 阅读本领域论文
  - 写作流程
  - 论文结构

- 投论文
  - 期刊&会议
  - 论文投稿是一场心理战
  - my point of view

一、论文写作

1. 如何进入本领域

◆ 工具：(快速上手写代码的能力)

- 熟练使用Python工具

  数据处理能力，基础算法
  sklearn,numpy,
  tf, pytroch, keras,
  dynet, tensor2tensor, fastAI, allenNLP, chainer …

- 其他：Linux，c++, java, CS basis.

一、论文写作

1. 如何进入本领域

◆ 内容上，掌握本领域的整体研究框架

- models, 在相关数据上跑过现有的demo， 有大致的原理理解:

machine learning models

分类模型，序列标注模型，主题模型
图模型（graph）
span prediction
transition

deep learning models

MLP, RecurrentNN, RecursiveNN, CNN,
seq2seq
many Gans,
reinforcement learning
Autoencoder: VAE
CapsuleNet
Transformer衍生

机器学习模型周边

统计模型的本质
数据理解，划分
对统计&分布的理解

一、论文写作

1. 如何进入本领域

◆ 内容上，掌握本领域的整体研究框架

- 以NLP tasks为例

- 分类：短文本分类，文档分类，情感分析
- 抽取：XX细粒度分析，NER，CWG，关系抽取，事件抽取，SRL，AMR
- 底层：dependency parsing, CFG, lexicon, semantic, 篇章
- 异步many-to-many: 翻译、阅读理解、自动文摘、seq2SQL
- others: transfer learning, multi-task, RL, multi-lingual,
- many new to come...

Domain adaptation
Entity linking
Information extraction
Named entity recognition
Part-of-speech tagging
story ending generation
peotry generation
text summarization
Word sense disambiguation
Relation prediction
Relationship extraction
Chinese word segmentation
punctuation，顿句
双关语检测
成语检测
第一人称检测
通顺度检测
幽默检测
隐喻检测 Metaphor detection
讽刺检测
word composition
Coreference resolution
Grammatical error correction
Simplification
Stance detection
Taxonomy learning
Temporal processing
paraphrase
sequence transduction (机器翻译，问答，文摘等等)
Stance analysis
Query auto-completion
行为链、事件链分析
Automatic essay scoring (AES)
Generating Classical Chinese Poems
Text normalization
Discourse segmentation
Entity Typing
sentiment modification
句子排序text-to-SQL
Hierarchical text classification

一、论文写作

1. 如何进入本领域

◆ 内容上，掌握本领域的整体研究框架

- 懂算法的原理以及周边相关

- 马尔科夫性
- 语言模型
- 反向传播中，loss是怎么影响参数学习的?
- 当提到数据分布时，到底在提什么?
- Logits
- Upper-bounds Inference
- Inductive Bias
- Exposure Bias

一、论文写作

Please notes:

- Learning by doing
- 体系学习，掌握本领域的整体研究框架
- 上面提到的模型，都是基本的，基础的，他们是不能直接拿去发文章的!
- 学术论文发表：需要理论上的创新度。

一、论文写作

2.阅读本领域论文和查新

◆ 本领域论文检索方式

- official：

  google scholar, bing, baidu scholar,
  glgoo, scihub
  dblp,
  Arxiv,
  ACL-antholody
  …

- Other source：

  知乎blog, 52NLP
  公众号：新智元，量子位，专知…
  …

一、论文写作

2.阅读本领域论文

◆ 怎么读论文

- 第一遍:

  标题，abstract，introduction，读一遍，是不是感兴趣的，大概做了个啥?

- 第二遍:

  模型是啥，看插图，
  跑的啥任务
  数据集是啥

- 第三遍

  methodology in details.
  contribution，motivation，innovation?
  为什么能中?

一、论文写作

2.阅读本领域论文

◆ 读论文，到底在读什么？

- 方法查新，
- 任务与数据集
- 创新点

- 本任务或者方法是否以及被人做过了

- 论文写法，行文风格，论文结构，
- 工作的构思，细节出奇制胜

一、论文写作

3.创作流程

- Step1.

  先确定工作内容，是否idea已经被做过了

  构思好：实验任务，大致模型方法，数据集，以及验证实验

  想清楚：contribution，motivation，innovation

- Step2.

  码代码，跑分

  记录数据

  写初稿

  增加实验设置

- Step3.

  完整稿子

  打磨语言表达

  检查漏洞

  重复以上

一、论文写作

4. 论文结构

- **Abstract**
- **Introduction**
- **Related Work**
- **Framework**
- **Experiment (settings, results, analysis)**
- **Conclusion**
- Reference

引文(Introduction)写法

- 论文的“门面”，必须写好
  - 引文一般包含的内容
    - 交代研究任务
      - 让读者知道你要做的任务
    - 阐述研究现状并总结不足
      - 给读者一些研究背景的铺垫，并且带出挑战以及难点
        - » 挑战如果非常重要，可以单独出一段
    - 提出解决的新思路
      - 引出论文的解决思路或者说idea
    - 给出新方案的设计
      - 较为详细地介绍idea的实现
    - 总结论文的贡献以及实验结论
      - 总结并且强调论文的贡献

引文(Introduction)写法

- 引文中的灵魂="逻辑"
  - 语法只是皮囊
- 常见错误逻辑
  - 因为模型A好使，所以用A做某任务（蓄意的科研）
  - 因为任务B没有人做，所以我做了（暴力的科研）
  - 之前的人做了什么工作，我做了什么工作，我的好（缺乏解释、对比）
  - 这个任务很难，这篇论文我们这样解决了它（缺乏过渡、解释）
  - 夸大自己模型的贡献、忽略别人的工作（Reviewer最反感的写法）

引文(Introduction)写法

- 引文中的“度”
  - 包装需要，但是要适宜

- 常见不合适的“度”
  - 随意给出一些非常主观的意见
    - 例如：CNN model is perfectly good at modeling xxx data
      - 加引用可能能缓解
  - 随意使用一些特别general的词汇
    - 例如：“knowledge”、“context”、“information”等
  - 随意夸大自己的模型、放大自己的贡献
    - 例如：
      - 之前的工作这也不行那也不行，就我的方法行
        - » 例子：Our model significantly improves over all previous methods
      - 建议从头到尾检查一下所有副词、形容词，去掉带有太多主观色彩的词汇，也要慎重使用一些程度强烈的词汇（如significantly）
      - 加一些限定词
        - » “on xxx task in terms of xxx metrics”
        - » “all previous methods”---“all the comparison methods”

引文(Introduction)写法

- 初学者写法-六句话扩充法
  - 第一句写任务介绍以及意义
  - 第二句概述研究现状以及主要的问题
  - 第三句写解决这些问题的研究挑战
  - 第四句写当前方法的主要出发点以及解决思路
  - 第五句写当前方法的主要技术方案
  - 第六句写总结、强调贡献

相关工作(Related work)写法

- 主要要求
  - 尽量覆盖所有相关的相关工作
  - 分类整理
  - 突出相关的地方
  - 强调不同的地方

- 常见错误
  - 简单罗列（平时读论文要做到分类总结）
    - 例如：A做了什么、B做了什么、C做了什么。。。
  - 没有说清楚区别和联系（平时读论文要做到分类总结）
  - 时态
    - 过去时
    - 现在完成时
    - 哪种都可以，但是不要混着来

模型部分(Method)写法

- 定义部分的写作
  - 介绍清楚所有术语
  - 给出所有符号的含义以及使用方式
  - 形式化地描述清楚任务

- Notation的例子
  - 部分问题
    - 一个符号多次使用，又代表不同意思
    - 全部使用未加粗的notation表示集合、矩阵等
    - 符号使用不按照习惯使用
    - 频繁使用一些单词的缩写用于notation (src, dest)
    - 频繁使用上下角标都存在的符号
    - 符号的数量非常多
    - \log, \exp \min …

模型部分(Method)写法

- 模型部分的写作
  - 逻辑很关键
  - 几种常见的逻辑
    - 总-分-式
    - 总-基础-增强式
  - 最后部分的讨论可能会给模型“添彩”
    - 正确性证明
    - 时间复杂度
    - 完整的算法流程
    - 参数汇总+学习算法
    - 与之前工作的区别（突出创新性）
    - 与之前工作的联系（增强泛化性）
    - 模型可扩展的地方（堵漏）

实验部分(Method)写法

- 一般流程
  - 数据集合、评测指标、评测流程
  - 对比方法
  - 主干实验分析
  - 模型细致分析
  - 定性实验

- 对比方法

Our baselines have a comprehensive coverage of the related models. To summarize, we categorize the baselines into eight groups shown in Table 2, according to the *task orientation*, *with/without KB* and *with/without neural models*.

Table 2: The categorization of the comparison methods.

Tasks	KB	Neural (*No*)	Neural (*Yes*)
General	*Yes*	–	CKE
	No	BPR	NCF
Sequential	*Yes*	–	RUM, GRU_F, **KSR**
	No	FPMC	GRU, GRU++

表格是一个总结对比方法的途径

- 主干实验
  - 要很清楚实验的目的
    - 对引文里面给出的贡献、发现或者结论的证明
    - 不要流水账一样介绍，要突出原因
    - 错误例子：A比B好、B比C好。。。
  - 有些异常结果要加以解释
    - 有些模型达不到原始论文的效果，要好好分析一下
  - 加上统计性显著检验
    - 确保提升是有效的
    - 有的时候是压死论文的最后一棵稻草

- 细致性分析实验
  - 检查contribution的来源
    - Ablation study
  - 组件内部调节
  - 参数调节
  - 数据调节

- 一般的流程
  - 避免探索式描写，要有核心驱动进行描写
  - 画一个好图：一图胜千言
  - 用好图的标题：各种符号、颜色以及整体场景的设置
    - 不要让文本描述和图的标题大部分一样，双方各有分工
  - 写一段清楚的描述
    - 首先写清楚目的
    - 接着写清楚当前例子的整体故事
    - 然后分解进入关键部分
    - 最后总结发现

推荐写作流程：

- 先写Related work，梳理清楚已有工作的不足，第一遍
- 然后改问题定义、模型，确保了解模型细节，第一遍
- 然后写Introduction，第一遍
- 然后写实验，第一遍
- Repeat“问题定义、模型、实验、相关工作”
- Repeat“Introduction、模型”
- 全文定稿
- 最后，abstract, conclusion，related work

二、投稿

1.期刊&会议

瞄准A类的目标去做

◆ 会议（CCF）

- ACL系列：ACL，EMNLP，NAACL，CoNLL, (+coling)
- AAAI, IJCAI， ECAI :
- 国内：NLPCC，CCL
- ！ICML, NIPS, ICLR: hardcore
- ！SIGIR, SIGKD, ...
- others:

  BIBM

  WWW

  ...

二、投稿

1.期刊&会议

◆ 期刊

周期长

- SCI系列
  相对容易些。
  - ✓ Neurocomputing
  - ✓ Information science
  - ✓ Knowledge base system

- CCF系列
  IEEE trans系列的期刊难度也是存在的，业内也是认可的。
  - ✓ Neural networks
  - ✓ TNNLS
  - ✓ TPAMI
  - ✓ TSLP
  - ✓ TASLP
  - ✓ JAIR

但千万别投 IEEE Access

二、投稿

2.论文

啥要给你的打高分？

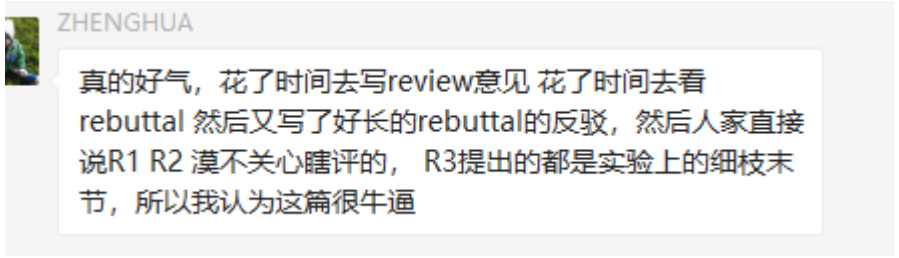
某知乎大V的回复真的让人无语

声称自己不是专家，还乱送满分，原因居然是因为自己的直觉？？？不是专家就说明缺少该领域的很多基本知识，我想象不出来哪来的自信会如此坚定的认为一个自己不了解的工作是完美的，deserve一个10分？？？就因为提供了代码？？？所以就不水了？？？

居然认为SPC不如小学生，SPC起码都是有一定的学术成就，大部分都是已经在学术圈摸爬滚打多年的prof们，就算不是专家，岂不是比你学术训练多多了？？？难道学术品位会比你差？？？

审paper，还要去判断作者的国籍，学校等个人信息？？这难道不违反双盲的宗旨吗？？双盲不就是让大家真正的面对工作实际的内容，而不是从其他信息去干扰自己的判断吗？？什么叫出自中国学校的概率是0？？难道中国那么多高校科研人员的工作都是灌水的？？没有优秀solid的paper吗？？？逆向民族主义者？？？感觉应该是欧洲或者以色列的科学家的工作？？？是在美国待久了被white wash了吗？？

就是因为每年这样随意给分的reviewer，才让顶会的公平性大打折扣，给分越来越随机。不尊重专业，不尊重权威，甚至有些racist，这样的reviewer 以后不审AAAI，对AAAI可能是个好事。



把握的，剩下的也不用过多忧患。

二、投稿

Strengths

1. The writing is decent
2. The technical aspect seem

Weaknesses

1. This idea is not particularly
   a little more explanation.
2. The technical exposition is
3. The design choices are ve
   network is just an attentior
   embedding. And for K-emc
   performing attention again
   for extracting features.
4. Even the core components
   network" are all off-the-she

Ove

Questions for t

Presentation Improvements

Reasons to accept

An improvement for general attention mechanism will be broadly beneficial for the NLP community. The idea is interesting, the experiments are convincing, and the improvement is obtained for different tasks and different kinds of attention mechanisms.

Reasons to reject

The improvement is not convincing for some tasks, i.e. one percent accuracy improvement does not show the strength of the proposed method sufficiently. And the proposed method seems over-complicated compared with those improvements.

Questions for the Author(s)

Is it possible to use a supervised learning method instead of reinforcement learning? It would be better if the authors can provide some reason for this choice since it seems to me that the difference between the defined reward function and the original loss function is a regularization term, which may be included to the final loss of the supervised training. In that way, the whole framework will be simpler and training could be faster.

Reasons to reject

- The model is not defined properly. The authors says this is "a novel Markov decision process" (L114), but as far as I read the paper, the model shown in the paper is not a Markov decision process. {\bf s}*t, the state of the time step t defined in Section 3.1.1, does not depend on the previous action a*{t-1}. All the states {{\bf s}*1, ..., {\bf s}*n} can be calculated without any action a_t. If the authors believe this is a Markov decision process

o solve, the reason should be explained.
ot clear. The function f(x)=x+0.01/x (L319) was
onale is not clear. I wonder if the constant value
e better at least to show experimental results
ward function in Equation (5) uses the value N
t is said "an accumulated threshold is set based on
setting is unclear and it is hard to reproduce the

el is low. In Table 5, the seq2seq model with DRGA
English-German. However, it is known that the
es 28.4 (Vaswani et al, 2017). It is better to run
hich are not weak.
roposed model improves the attention. In Section
n mechanism is mentioned (L079-082), but it is not
s the attention quality. Some examples of the
6 and 7. However, seeing a few successful results
entions were improved and worsened. In NMT,
EU has been calculated in studies to improve the
entions for Neural Machine Translation; Liu et al.:
vised Attention). It will be better to evaluate with

ise.
tion since \alpha_t is a scalar.
h^s} are equal, and the same symbol should be used for avoiding confusion.

- (L274) s_i -> s_t
- (L296) \alpha* -> {\bf \alpha}*
- (L358, 359) {\bf w}*i -> {{\bf w}*i}_{i=1}^n, {\bf h}*i -> {{\bf h}*i}_{i=1}^n

二、投稿

other:

1. [Summary] Please summarize the main claims/contributions of the paper in your own words.
In this paper, authors try to employ RL method to refine attention mechisame to make sure that the attention weights are more related to the task. They show that this method could improve a wide range of NLP applications and improve the learning efficiency.

2. [Relevance] Is this paper relevant to an AI audience?
Likely to be of interest to a large proportion of the community

3. [Significance] Are the results significant?
Highly significant

4. [Novelty] Are the problems or approaches novel?
Somewhat novel or somewhat incremental

5. [Soundness] Is the paper technically sound?
Technically sound

6. [Evaluation] Are claims well-supported by theoretical analysis or experimental results?
Very convincing

7. [Clarity] Is the paper well-organized and clearly written?
Good

8. [Detailed Comments] Please elaborate on your assessments and provide constructive feedback.
1. They incorporate a policy network to refine attention weights automatic.
2. This method is effective in many tasks.
3. Analyzing is reasonable and intuitive.

9. [QUESTIONS FOR THE AUTHORS] Please provide questions for authors to address during the author feedback period.
`six challenges for neural machine translation` and `Neural machine translation with adequacy-oriented learning` are also related to this topic.

---

10. [OVERALL SCORE]
8 - Accept (Top 50% accepted papers (est.))

7. [Clarity] Is the paper well-organized and clearly written?
Satisfactory

8. [Detailed Comments] Please elaborate on your assessments and provide constructive feedback.
Pros:
1. Extensive studies are conducted on various model structures

Cons:
1. This formulation modifies attention based on the current hidden state, which is available to the attention module as well, then why does this method provide better results? Is it due to the added parameters or due to the residual modeling? A reasonable baseline would be doing residual modeling of attention and add outputs from an MLP to the original attention weights and train in the normal fashion.
2. I don't understand why the policy gradient is needed. Is there any randomness of the attention modifications? It's true that we cannot directly backpropagate through the second term of the reward, but 1) that term looks weird (see below), and 2) we can do a continuous version of that too.
3. How to set the additional hyperparameters beta, L, threshold, and gamma? Grid search using validation performance seems to be very expensive. Besides, is the second term of the reward necessary? What's the intuition?

9. [QUESTIONS FOR THE AUTHORS] Please provide questions for authors to address during the author feedback period.
See the above comments. Besides, in Figure 3 and Table 2, what's the task?

10. [OVERALL SCORE]
4 - Reject

二、投稿

2. my point of view

- **工作量一定要充分！**

  **动机，要自圆其说。

  ***方法论模型硬货

  *细节非常丰富

  ***实验非常充分，baseline充分查新，各个方面角度都得到验证

  *没有低级错误（书写，数字错误，引用格式），防止被抓小辫子

二、投稿

2. my point of view

- **想上论文，最好得跟上state-of-the-art,**

  比如，别人的baseline加上了elmo, bert，验证了效果，则有新的突破点

二、投稿

2. my point of view

- **快2020年了，早已过了上上新模型就能发顶会系列的时期了，岂能随便在现有的模型上小修小改跑个任务就完事?**

  想要搞定reviewer的胃口，硬货部分要充分明显。
  理论创新充分！模型改进明显；引入了新的linguistic上的理论点，等等。
  总归一点：一定要让人感到眼前一亮。

二、投稿

2. my point of view

- **提升代码实现能力**

  随意复现、修改、实现模型结构

  LSTM cell level modification

  ...

二、投稿

2. my point of view

- **选一个对的方向/任务, 也非常关键!**

  简单分类，简单入门任务？No!

  烂大街任务：情感分析，关系抽取

  ...

二、投稿

2. my point of view

- **一定要刷论文集，整套整套地。**

  比如 ACL、EMNLP、AAAI 2020论文集。

  查新

  思考别人为什么会中。

QA