Doing research

These mindmaps are the supplementary materials for "A brief and amusing introduction to doing research"

Possible choice for a research topics

- Propose a new technique: read the conclusion & future works of papers, read surveys, papers on open problems (careful about publication year),...
- Study properties or effects of a known method on something
- Compare known methods, possibly for specific types of data/applications, especially new techniques
- Survey papers: missing data > time series > imbalanced
- Read survey articles to see what people have done in this field (ACM survey)
- Generalization: consider the more general case of a problem
- Application of a known technique to biology, finance,...
- Problems for special cases: is it possible to find a more efficient way to speed up/ compute for this special case?
- Write about the properties of an existing method
- Compare different methods: why choose to compare these methods, has anyone done such a survey? Comparable for a particular data type
- Do a survey paper

Choosing a topic:

- Interests
- Difficulty of the problem? Starter? Can we start with a simpler problem?
- Number of people in the project
- Infrastructure & data sizes
- Target revenue

Surveys

- ACM Computing Surveys
- Artificial Intelligence Review

Experiments

- Keep different versions of the codes
- Keep the sides data for the experiments: some reviewers require mean $\pm$ standard deviation, graphs,...
- output of an experiments should be easy to view for better analysis
  - example python output result table

```
[(       ADASYN  kmeanSmt     SmtNN    SVMSmt   InsHard     OsImp
 SVM  0.904444  0.903015  0.870502  0.895956  0.745422   0.90391
 LR     0.56098  0.578764  0.552504  0.581332  0.621738  0.636731
 DCT  0.955189  0.953642  0.911244  0.954357  0.812707  0.963621,
         ADASYN  kmeanSmt     SmtNN    SVMSmt   InsHard     OsImp
 SVM  0.899277  0.897688  0.861994  0.890173  0.734971  0.903324
 LR   0.513728  0.531358  0.492775  0.537861  0.606358  0.685405
 DCT  0.955202  0.953468  0.907659  0.954335  0.792775  0.963584),
 (       ADASYN  kmeanSmt  SmtNN  SVMSmt  InsHard  OsImp
 SVM    0.009     0.013  0.010   0.013    0.023  0.017
 LR     0.014     0.017  0.013   0.022    0.017  0.012
 DCT    0.009     0.009  0.015   0.013    0.023  0.007,
       ADASYN  kmeanSmt  SmtNN  SVMSmt  InsHard  OsImp
 SVM    0.009     0.015  0.011   0.014    0.029  0.017
 LR     0.017     0.019  0.016   0.031    0.025  0.013
 DCT    0.009     0.009  0.017   0.013    0.032  0.007)]
```

  - use plot to easily detect trends, patterns in the results
- comment the codes carefully and have multiple people to look at the codes
- simulation on toy examples can be important: $X\sim N(0,1)$

Writing

Overleaf & Writing

- recommended book: Writing for computer science - Zobel
- Avoid deleting s.o. else writing unless you're ... their boss. Comment it out
- references `\ref` should be named for some reason: `\ref{sec-intro}, \ref{sec-related-works}, \ref{sec-concl}`, `\ref{tab-datasets}, \ref{tab-mnist-result, \ref{tab-digits-res}`, `\ref{fig-histogram}, \ref{fig-lesion}`
- the best first: best result tables first, the not so good for later; advantage first, drawbacks later
- consistent writing: use Table/table, figure/ Figure
- Use interpretable notations, avoid repetitive notations, follow standard rules
- As you insert a lemma from another paper/ book, you need to make sure that the notations are consistent with the current writing style, instead of pure copy and paste. For example, if the statement you want to insert uses A' for matrix transpose, and your paper use A^T for the transpose instead, you need to switch the notation of the statement you want to insert to A^T

Abstract

- Very important because the summary is the first thing the reader will read
- Give reasons for recommending this good method to write this article
- State the main findings of the paper
- whether the experimental results are good or not, in what respects (accuracy, MSE, speed, ...)

introduction

- Highlight why recommend this approach
- Outline what progress has been made by the research community?
- State the contribution of the article
- Organizational structure for the rest of the paper

Related works

- Advances in the field, especially the most recent ones (Groups, write one paragraph at a time)
- Strengths or weaknesses of relevant articles
- The proposed algorithm is based on what methods?

methodology

- Presenting the introduced spell
- If too many symbols are used, a symbol table should be used
- Choose symbols appropriately (traditionally, in a sense)
- Break into small sentences if possible
- Divide into small items

experiments

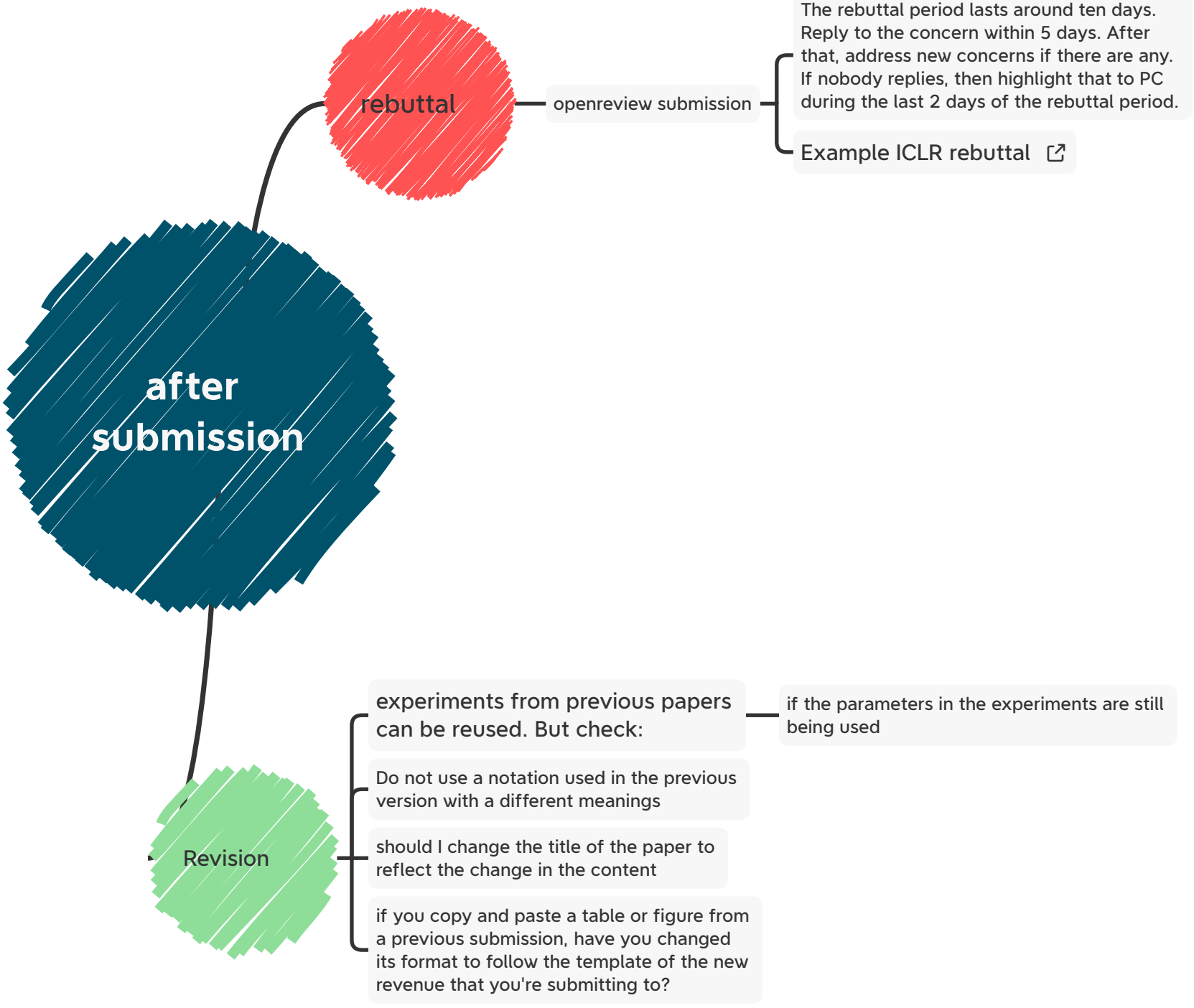
- Compare with: recently developed methods & popular methods
- Compare on data sets of different sizes & different data types
- Runtime (on big data is fine)
- Where are the results reported?
- Begin the paragraph with the strength of your method. Then talk about the weak point.
- One main point for each paragraph
- Give specific proof
- Chart
- My weakness is not that weak
- Abnormal: explain why (including methods of other papers)
- Run a step-by-step/partial test test & check which step/part is the problem

Conclusion

- Summarize the main idea, strengths and limitations of the algorithm
- Issues to be studied in the future

submission checklist

- grammar & typos in title
- - If you adjusted your algorithm, have you updated that algorithm and its analysis in the methodology section? have you updated examples/ experiments analysis?
- - Is everything consistent? Ex: missing data at 20%-60% at tables but 20%-80% at a graph
- - Abstract: recheck if the important remarks that you found after running experiments are there yet
- notations
  - the notations has been clearly defined
    - if there are many notations: create a table of notations. The table can be placed in the main content or the appendix
  - same notations with different meanings
- experiments
  - - If you update the experiment results, have you updated have corresponding analysis?
  - - If you added one more experiment on a dataset, and you copy and paste the table formatting, have you updated the table's label?
- example
  - I created an example dataset D1 in the introduction. Then, I used it to illustrate algorithm A in example 1. Then, I updated dataset D1, but forgot to update example 1
- template format is correct, including tables, graphs, etc.


after submission
rebuttal
openreview submission
The rebuttal period lasts around ten days. Reply to the concern within 5 days. After that, address new concerns if there are any. If nobody replies, then highlight that to PC during the last 2 days of the rebuttal period.
Example ICLR rebuttal
Revision
experiments from previous papers can be reused. But check:
if the parameters in the experiments are still being used
Do not use a notation used in the previous version with a different meanings
should I change the title of the paper to reflect the change in the content
if you copy and paste a table or figure from a previous submission, have you changed its format to follow the template of the new revenue that you're submitting to?


Writing presentation slides and presenting your work
you want to persuade people to read and use your research
write maximum n slides for a talk of n minutes, not including the thank you slide
Highlight the best or most important points
Don't cite a lot like in your research paper
Theorem/equations: tell the reader what they mean or imply when you give the talk but do not write it in the slides
avoid unnecessary theorems/ theoretical results
Can highlight major points from experiments results instead of experiment details
Talk about future works so that people know that they should follow your work or they can collaborate with you