NYU

Section 12
How to publish your work

Xiang Pan

11/27/2023

Project

Heilmeier's Criteria

- What are you trying to do? **Articulate your objectives using absolutely no jargon.**
- How is it done today, and what are the limits of current practice?
- What is new in your approach and why do you think it will be successful?
- Who cares? If you are successful, what difference will it make?
- What are the risks?
- How much will it cost?
- How long will it take?
- What are the mid-term and final “exams” to check for success?

Writing

Writing

- If you want to publish an influential paper in NLP, the engineering is only half the work.
- The general pipeline:
  - How to find and understand related work on your problem
  - How to design effective experiments and analyze their results
  - How to stay out of ethical trouble
  - **How to write and publish your work**

Three ways to organize your ideas (Shieber)

- Continental style: “in which one states the solution with as little introduction or motivation as possible, sometimes not even saying what the problem was” [. . .] “Readers will have no clue as to whether you are right or not without incredible efforts in close reading of the paper, but at least they’ll think you’re a genius.”
- Historical style: “a whole history of false starts, wrong attempts, near misses, redefinitions of the problem.” [. . .] “This is much better, because a careful reader can probably follow the line of reasoning that the author went through, and use this as motivation. But the reader will probably think you are a bit addle-headed.”
- Rational reconstruction: “You don’t present the actual history that you went through, but rather an idealized history that perfectly motivates each step in the solution.” [. . .] “The goal in pursuing the rational reconstruction style is not to convince the reader that you are brilliant (or addle-headed for that matter) but that **your solution is trivial**. It takes a certain strength of character to take that as one’s goal.”

Abstract

- Important for creating a first impression, reviewer bidding, and reviewer assigning.
- A general structure:
  1. The opening is a broad overview — a glimpse at the central problem.
  2. The middle take concepts mentioned in the opening and elaborates upon them, probably by connecting with specific experiments and results from the paper.
  3. The close establishes links between your proposal and broader theoretical concerns, so that the reviewer has fresh in her mind an answer to the question "Does the abstract offer a substantive and original proposal".

Figure One (The main idea)

It should be possible to understand this figure without reading the rest of the paper.

Abstract

Semantic word spaces have been very useful but cannot express the meaning of longer phrases in a principled way. Further progress towards understanding compositionality in tasks such as sentiment detection requires richer supervised training and evaluation resources and more powerful models of composition. To remedy this, we introduce a Sentiment Treebank. It includes fine grained sentiment labels for 215,154 phrases in the parse trees of 11,855 sentences and presents new challenges for sentiment compositionality. To address them, we introduce the Recursive Neural Tensor Network. When trained on the new treebank, this model outperforms all previous methods on several met-

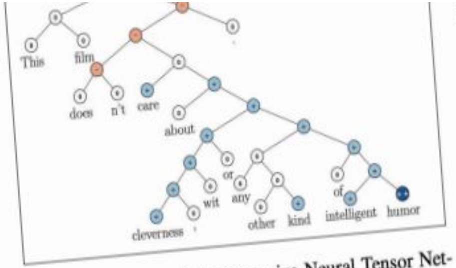


Figure 1: Example of the Recursive Neural Tensor Network accurately predicting 5 sentiment classes, very negative to very positive (– –, –, 0, +, + +), at every node of a parse tree and capturing the negation and its scope in this sentence.

NYU

Formatting

- Latex
- Style File
  - Do not change!
- Citation
  - BibTex
  - Bib
    - Google Scholar -> Last Choice!
    - http://aclweb.org/anthology/P/P17/P17-1060.pdf
      http://aclweb.org/anthology/P/P17/P17-1060.bib
  - Zotero

Plot

- https://www.mathcha.io/
- Tikz
    - Anything!
- ChatGPT
    - Figure -> Latex Code
- BERT-Specific
    - https://docs.google.com/presentation/d/1HMng1RWuY1molsGamwIpRnxNZNEqEPRVODcXpm-pX4c/edit#slide=id.g8e6447f66a_2_1632

Checklist [Saunders et al. (2009)]

- Is there a clear structure?
- Is there a clear storyline?
- Does your abstract reflect accurately the whole content of the report?
- Does your introduction state clearly the research question(s) and objectives?
- Does your literature review inform the later content of the report?
- Are your methods clearly explained?
- Have you made a clear distinction between findings and conclusions in the relevant chapters?
- Have you checked all your references and presented these in the required manner?

Checklist [Saunders et al. (2009)]

- Is there any text material that should be in the appendices or vice versa?
- Does your title reflect accurately your content?
- Have you divided up your text throughout with suitable headings?
- Does each chapter have a preview and a summary?
- Are you happy that your writing is clear, simple and direct?
- Have you eliminated all jargon?
- Have you eliminated all unnecessary quotations?
- Have you checked spelling and grammar?
- Have you checked for assumptions about gender?
- Is your report in a format that will be acceptable to the assessing body?

Publishing

- If successful, finished paper should be the foundation of a publishable research paper.
- Major conferences have higher expectations for ambition and for thoroughness of analysis
- Build on what you've found until you have a substantial result that you're confident about, and submit them!

Why paper get rejected

Overclaiming

The easiest way to get a paper rejected (or given a low grade):

- Saying something that isn’t true.

Almost as easy- >Saying something that’s true without sufficient evidence.

- The paper makes a concrete claim that isn't backed up by appropriate citations or direct evidence.
- Reviewers may allow for some overclaiming when describing related work and background, if they trust that you'll fix it. Any overclaiming about your own contributions will result in a rejection.

Why paper get rejected

Overclaiming

- The paper makes a concrete claim that isn't backed up by appropriate citations or direct evidence.

Examples:

- In an introduction: "Researchers have long struggled to do XYZ." (Do you have evidence that people actually worked on this?)
- In an introduction: "Unlike current ML models, humans do XYZ by reasoning about concepts like ABC." (Have cognitive scientists really concluded this?)
- In related work or methods: "BiLSTMs are the best approach to task ABC (XYZ et al. 2018)." (Did XYZ really show that it's the best approach? Is this still true now?)
- In an abstract or conclusion: "We show that our system beats BERT." (Did you run a fair comparison with BERT?)

Why paper get rejected

Overclaiming

- The paper makes a concrete claim that isn't backed up by appropriate citations or direct evidence.

Examples:

- In an introduction: "Researchers have long struggled to do XYZ." (Do you have evidence that people actually worked on this?)
- In an introduction: "Unlike current ML models, humans do XYZ by reasoning about concepts like ABC." (Have cognitive scientists really concluded this?)
- In related work or methods: "BiLSTMs are the best approach to task ABC (XYZ et al. 2018)." (Did XYZ really show that it's the best approach? Is this still true now?)
- In an abstract or conclusion: "We show that our system beats BERT." (Did you run a fair comparison with BERT?)

Why paper get rejected

Overclaiming (commen)

- The paper makes a concrete claim that isn't backed up by appropriate citations or direct evidence.

Examples:

- In an introduction: "Researchers have long struggled to do XYZ." (Do you have evidence that people actually worked on this?)
- In an introduction: "Unlike current ML models, humans do XYZ by reasoning about concepts like ABC." (Have cognitive scientists really concluded this?)
- In related work or methods: "BiLSTMs are the best approach to task ABC (XYZ et al. 2018)." (Did XYZ really show that it's the best approach? Is this still true now?)
- In an abstract or conclusion: "We show that our system beats BERT." (Did you run a fair comparison with BERT?)

Why paper get rejected

Methodological issues (common, related):

- Your methods are almost sufficient to make some interesting claims, but there's a crucial flaw that makes the results hard to interpret.

Examples:

- Did you tune your baseline correctly?
- If you're working with pretrained models, does the model's tokenization and vocabulary make sense for your task?
    - Digital Task
- Did you use the right metric for the claim you're making?

Why paper get rejected

Motivation (somewhat common, related):
You answer a research question, but don't explain why a reasonable person would ask that question.
Examples:
- If you're using NLP for a problem for the first time, could someone actually use NLP for your problem in the real world without hitting ethical/legal/logistical issues?
- If you're trying to improve the performance of some system that isn't the state of the art, is there some reason that a user would want to use that system?

Why paper get rejected

Motivation (somewhat common, related):
You answer a research question, but don't explain why a reasonable person would ask that question.
Examples:
- If you're using NLP for a problem for the first time, could someone actually use NLP for your problem in the real world without hitting ethical/legal/logistical issues?
- If you're trying to improve the performance of some system that isn't the state of the art, is there some reason that a user would want to use that system?

Why paper get rejected

Motivation (somewhat common, related):
You answer a research question, but don't explain why a reasonable person would ask that question.
Examples:

- If you're using NLP for a problem for the first time, could someone actually use NLP for your problem in the real world without hitting ethical/legal/logistical issues?
- If you're trying to improve the performance of some system that isn't the state of the art, is there some reason that a user would want to use that system?

Why paper get rejected

Novelty (somewhat common)

- Did someone else already answer this question? If so, did you explain why it was necessary to revisit the question?

Impact (less common):

- Will at least a few dozen people find this paper relevant to their own work in the future?

Fit (**less common** for larger conferences, though relevant to workshops):

- Do members of your intended audience tend to read papers that are published here? Do your scientific peers tend to review for this venue?

Sharing Your Code!

Standard practice

- Make a GitHub repo for all of your code and saved model files for your best runs.
- Prepare a readme with instructions on:
  - How to access any necessary data
  - How to install and configure your code
  - How to retrain your model to reproduce your experiments
  - How to run your trained model on new data
- All of this should be sufficient to reproduce all of your results without contacting you.
- Include a link to the repo in your paper.
- Best practice is to anonymize the whole repo during blind peer review, but many people just censor the URL and only include it after review.
  - https://anonymous.4open.science

Presentation

The lightning Talk

Tips

- Don't put anything on your slides that you won't talk about.
- Don't talk faster than you do normally.
- If the listener wants more information, they can ask questions…
  - …but if they can't keep up with the pace of the talk, they'll just stop paying attention.
- Be honest about your conclusions and limitations.
- It's okay to present a slightly simplified version of your idea, as long as you're not misleading the audience.
- Only use technical terms if you have time to explain them!
- Practice, simplify, and practice again!
  - For a short talk, 25+ rehearsals is normal.

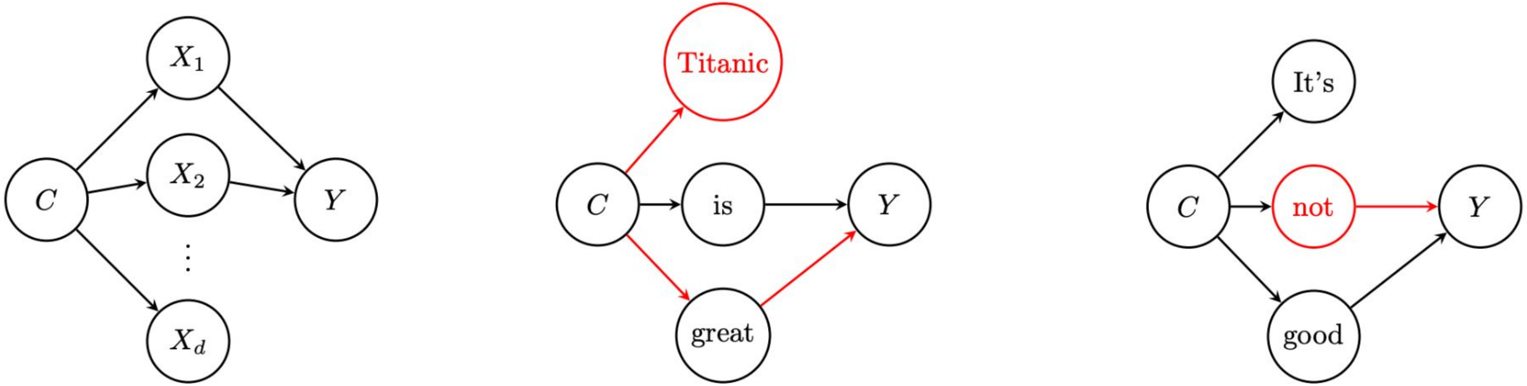
Example (Are all spurious features in natural language alike? an analysis through a causal lens. Nitish Joshi, Xiang Pan, He He. 2022 EMNLP)

Irrelevant features
Speilberg's new film is brilliant. ⟶ Positive
______'s new film is brilliant. ⟶ Positive
Necessary features
The differential compounds to a hefty sum over time.
The differential will not grow ⟶ Contradiction
The differential will ____ grow ⟶ ?

Table 1: Difference between two spurious features: (a) the director name can be replaced without affecting the sentiment prediction; (b) the negation word is necessary as it is not possible to determine the label without it.

NYU

Example (Are all spurious features in natural language alike? an analysis through a causal lens)



(a) Data generating model. (b) Type 1 dependence. (c) Type 2 dependence.

Figure 1: Causal models for text classification. (a) C is the common cause of words in the input. Each word X_i may be causally influence Y. (b) Y (sentiment label) and X_i ("Titanic") are dependent because of the confounder C (indicated by the red path). (c) Y (sentiment label) and X_i ("not") are dependent because of a causal relation.

Example (Are all spurious features in natural language alike? an analysis through a causal lens)

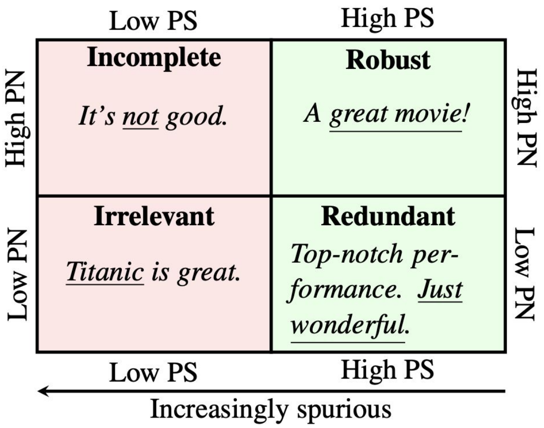


Figure 2: Categorization of features based on their PN and PS. Spurious features have low PS. Among them, the high PN ones are part of the features needed for prediction but they alone are not sufficient; and the low PN ones are irrelevant to prediction.

Example (Are all spurious features in natural language alike? an analysis through a causal lens)

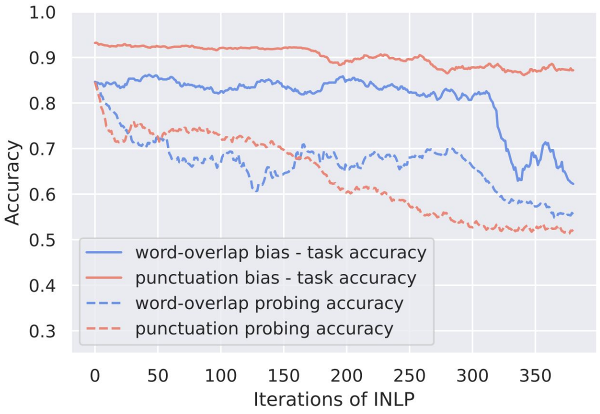


Figure 4: Extractability (probing accuracy) of the spurious feature (shown in dashed lines) and the task accuracy (shown in solid lines) as a function of iterations in INLP. For high PN features (word-overlap), its removal (decreasing probing accuracy) is accompanied by large drop in the task accuracy.

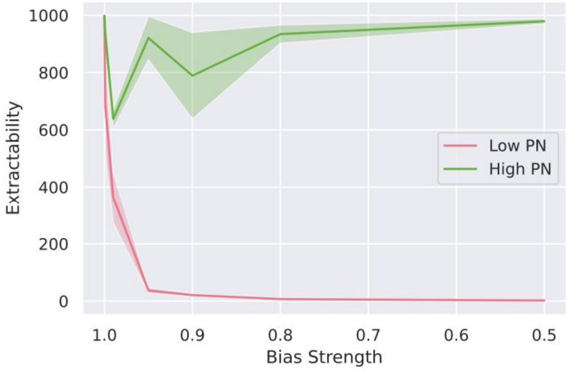


Figure 5: Extractability (compression) of the spurious feature as a function of bias strength on the synthetic data. The high PN feature is easily extractable regardless of its correlation with the label, whereas the low PN feature becomes less extractable when the bias strength drops.

Reference

- Bowman NLU 2020 Spring
- CMLS Presentation