Literature Review

3 April 2024
This presentation discusses research on using large language models (LLMs) to build explainable classifiers. It will show off work from my PhD students and collaborators on several recent research directions:

- Image classification with explainable features (arxiv.org/abs/2211.11158)
- Text classification with explainable features (arxiv.org/abs/2305.12696 and arxiv.org/abs/2310.19660)
- The importance of faithfulness in explanations (arxiv.org/abs/2209.11326)
- A faithful “chain of thought” LLM reasoner that produces code in its explanations (arxiv.org/abs/2301.13379)
Peer review
Dear Reviewer 2: Go F’ Yourself

David A. M. Peterson, Iowa State University

Objective. The objective of this study was to empirically test the wide belief that Reviewer #2 is a uniquely poor reviewer. Methods. The test involved analyzing the reviewer database from Political Behavior. There are two main tests. First, the reviewer’s categorical evaluation of the manuscript was compared by reviewer number. Second, the data were analyzed to test if Reviewer #2 was disproportionately likely to be more than one category below the mean of the other reviewers of the manuscript. Results. There is no evidence that Reviewer #2 is either more negative about the manuscript or out of line with the other reviewers. There is, however, evidence that Reviewer #3 is more likely to be more than one category below the other reviewers. Conclusions. Reviewer #2 is not the problem. Reviewer #3 is. In fact, he is such a bad actor that he even gets the unwitting Reviewer #2 blamed for his bad behavior.

Anyone who has ever submitted a paper to a peer-reviewed outlet knows the reviewers can, occasionally, be unpleasant. While rejection always stings, the belief that a reviewer has either completely missed the point of the manuscript, been overtly hostile in his or her review, or simply held the author to an impossible standard is vexing. The source of this frustration has seemingly become personified in the identity of a single person—Reviewer 2. He (and it is always assumed to be a he) is embodiment of all that we hate about other scholars. Reviewer 2 is dismissive of other people’s work, lazy, belligerent, and smug.

The purpose of this article is to test a very specific claim about Reviewer 2: he is the reviewer who holds us back. Using the database of reviewer responses from four years of Political Behavior, I empirically test if reviewers who are assigned number 2 are systematically more negative and more likely to be out of line from the other reviews a manuscript received. I assess this hypothesis in two ways. First, I compare the ordinal score each reviewer assigns the manuscript. Second, I develop an ordinal measure of “being Reviewer 2.”
ACL Rolling Review

A new initiative of the Association for Computational Linguistics

Review Form

Paper Summary

Describe what this paper is about. This should help action editors and area chairs to understand the topic of the work and highlight any possible misunderstandings.

Summary of Strengths

What are the major reasons to publish this paper at a selective *ACL venue? These could include novel and useful methodology, insightful empirical results or theoretical analysis, clear organization of related literature, or any other reason why interested readers of *ACL papers may find the paper useful.

Summary of Weaknesses

What are the concerns that you have about the paper that would cause you to favor prioritizing other high-quality papers that are also under consideration for publication? These could include concerns about correctness of the results or argumentation, limited perceived impact of the methods or findings (note that impact can be significant both in broad or in narrow sub-fields), lack of clarity in
Most scientists regarded the new streamlined peer-review process as ‘quite an improvement.’
Writing a good literature review
We write a literature review (or “Related Work” section) to:

- Acquaint the reader with the field of study
- Establish relevance
- Reveal literature gaps
Good literature reviews

Organize paragraphs around ideas
Describe studies in your own voice
Focus on a research question
Contain clear topical sentences
A literature review needs to be clearly organized.

This is a bit like a jigsaw puzzle – when you read papers, you’re finding the pieces, and then you need to figure out how they fit together.

If you have trouble finding some pieces, that may suggest that there is an important gap in the literature, which can be a great opportunity!
Common issue: No synthesis!

Do not:

- Write each paragraph about an individual study
- Try to report all the details from the paper

Rather, you’re describing the literature to set the stage for your own work.

- Have people looked at this problem?
- This data?
- What did they try?
- How did it work out?
Poor literature reviews contain

- Lots of in-depth details about the studies
- Too much jargon
- Topic sentences that discuss specific authors/studies
- References without citations
- Very long paragraphs
Peer review exercise
Pair off with someone near you who isn’t in your group.

If you have a print copy of your literature review, give it to the other person.

If you have an electronic copy, share that.
I recommend answering the following questions in a Google Doc, but it’s up to you.

Be prepared to share your written responses.
1. Is there a clear organization to the survey? If so, what is it?
   E.g.,
   chronological
   by problem
   by method
   etc.

2. Do you notice obvious gaps in the work described? What are they? Are they explicitly addressed?
   E.g., who introduced the problem? What is the best performing system?

3. How are the ideas in the lit. review being connected to the proposed study or research question?
   E.g., “While these papers show that topic modeling has been widely applied to problems in the humanities, to the best of our knowledge, no one has applied dimensionality reduction to uncover latent similarities in Lovecraftian horrors.”

4. Any other comments? How could this survey be more effective?
Share your answers with the other person.
Acknowledgments

The class incorporates material from:

Ruth Szpunar