Paper: Troubling Trends in Machine Learning Scholarship

Summary

In this paper, the author focusses on four trending patterns in ML scholarships that can make them flawed which can mislead the public and stymie future research by compromising ML's intellectual foundations. They hypothesize some causes for this trend which include the rapid expansion of the community, the consequent thinness of the reviewer pool, and the often-misaligned incentives between scholarships and short-term measures of success. Finally, they offer some suggestions that will help both experienced and inexperienced ML authors from avoiding such trends.

The first trend that authors stress is explanation vs. speculation. The authors explain the role of speculation as a means for authors to impart intuitions that may not yet withstand the full weight of scientific scrutiny. However, papers often offer speculation in the guise of explanation, which can mislead other researchers using that paper for their research work. The author stresses that it is important to distinguish speculation from facts.

The second trend is the failure to identify the sources of empirical gains. This can potentially generate a false impression that the authors did a lot of work when the improvements might have come due to a small change or hyperparameter tuning. Therefore, the authors mention that it is important to perform ablation studies.

The third trend the authors talk about is mathiness, which is a concept proposed by economist Paul Romer that describes the idea of using both natural language and mathematics to explain concepts but failing to provide tight links between statements and symbols. Because of Mathiness, the clarity of paper may suffer, and the reader of the paper may not be able to grasp the essential concepts of the paper.

The final trend is the misuse of language. The authors identify three types of language misuse in ML Scholarships: suggestive definitions, overloaded terminology, and suitcase words. Suggestive definitions are the terms with suggestive colloquial meaning. They sneak in connotations without the need to argue for them. They often manifest in anthropomorphic characterizations of tasks and techniques. Overloaded terminologies refer to the inappropriate or contradictory use of a technical term that already holds a very precise meaning in the literature. For example, the misuse of the term "deconvolution". Suitcase words are the terms that may hold many different meanings. Such words may create confusion among readers with different backgrounds.

The authors give three potential causes casual factors to such trends in ML scholarships. First, they claim that the authors may feel that it is okay to make weak arguments because they have some strong results, which they call complacency in the face of progress. The second cause is the growing pains. ML is a vastly expanding field and with that, a lot of inexperienced researches enter the field who are more susceptible to fall for such trends. The final cause is misaligned incentives. With popularity, ML practices face increasing pressure from media and startup investors to come up with some groundbreaking results. Such incentives may lead the researchers to weakly formulated paper.

Finally, the authors offer some suggestions to both experienced and inexperienced ML authors. They encourage the authors to identify "what worked" and provide clear details explaining "why it worked" and not just "how well it worked". They also stress the importance of error analysis, ablation studies, and robustness checks.

Strengths

- The authors have done a good job of structuring the paper into various sections and explaining each section clearly. There is also a nice flow between sections.

Weaknesses

- I wish they have also talked a little bit more about the visual aspect of the paper. Increasingly, visuals are becoming a core part of the paper.

Discussions

- How influential was this paper? Were they able to bring some changes in the trend of ML literature?
- Does ML researchers who are affiliated with a university has pressure from University to publish frequently as well?
- Do the suggestions that the authors offer in this paper be applied to various sectors that utilize machine learning? For example, do the same rule applies to an ML paper with a medicine background vs. finance background?