**Sentiment Analysis on the PeerRead Dataset**

**== Motivation ==**

In this project you will explore sentiment analysis on the paper review dataset PeerRead. The OpenReview platform was introduced in 2012-2013. The beginning of OpenReview was discussed in [a blog post from Yann Lecun](http://yann.lecun.com/ex/pamphlets/publishing-models.html):

*“Basically, this system plays the role of a "stock exchange", an open market of ideas”*

Existing peer review process is basically like this: each paper will be assigned with 3-4 reviewers. Each reviewer independently assign scores to assess the value of the paper. The overall score shows individual reviewer’s decision on whether the paper should be accepted, which is usually one of the following 6 options: strongly accept, accept, weak accept, weak reject, reject, strongly reject. After the reviews of individual reviewers, the area chair (a senior reviewer) will also read the paper, assess the value and make the final decision by considering the opinions of individual reviewers.

OpenReview also leverages the individual reviewer -> area chair process, the difference from existing review process are: (1) your paper is published as soon as they are submitted, anyone can see it; (2) all the reviews are published and anyone can see it; (3) anyone can review the paper by commenting on the paper, but such unofficial reviews do not impact the final decision (or do they?)

The OpenReview platform provides a dataset that could support NLP researches such as sentiment analysis and text summarization. It may also help with building tools such as quality assessment of reviews, detecting fake reviews, and supporting area chairs to prioritize their decision-making tasks.

Overall there is a high chance of publication in this direction. If you are interested please contact me (Xueqing).

**== Sentiment Analysis on PeerRead ==**

The task of sentiment analysis/classification is to decide whether a sentence shows a positive, negative, or neutral sentiment. Sentiment analysis research depends on the domain being studies. For instance, movie reviews are different from e-Commerce product reviews.

We can study sentiment analysis on the PeerRead dataset. Given a paper review (text), can we predict the score assigned by the reviewer using machine learning?

Sentiment analysis study also focuses on interpretable models, such as the aspect model [2]. The overall opinion towards a hotel reservation may be decomposed into multiple aspects including the location, service, price, and cleanness. Similarly, the overall opinion towards a paper review may be decomposed into the paper’s novelty, soundness, presentation, and experimental results. Can we predict what scores a paper receives in each aspect? How does the final review relate to the scores in each aspect?

**== PeerRead Dataset ==**

<https://github.com/allenai/PeerRead>

The dataset should contain all the review texts from the ICLR 2017 papers. If it doesn’t contain some data, it is also fairly each to crawl all the data from the website. I have a crawler in Python that I will try to find it and share it later.

Each paper should contain the following data components:

1. The paper text
2. The individual review text, scores, and confidence level
3. The area chair’s decision text, and the decision
4. Other unofficial review texts
5. The aspect scores assigned by their human annotators

**== Project Ideas ==**

Explore sentiment analysis on the PeerRead dataset. You can also explore other topics than sentiment analysis. It can be any of the following directions, or combining any two directions:

1. Predicting individual reviewer’s decision:
   1. Given the text of a review, can you predict the aspect-level sentiment in the review? How supportive is the reviewer towards the soundness of the paper? Do the three reviews often share the same or different opinions towards the soundness (or novelty, presentation) of a paper? What are some indicative words for sentiment? E.g., “yummy” or “such” indicates positive/negative opinion towards food, what are the corresponding words in the peer review dataset?
2. Predicting area chair’s final decision:
   1. Given the texts from each individual reviewer, can you predict the final decision from the area chair? How does it relate to the individual reviews? Is there any cases where you find the overall decision differ from individual reviewers’ opinion? If so why? The area chair has to review hundreds of papers, could you help reduce their efforts by prioritizing the papers that are the most uncertain given the individual reviews?
3. Comparison between two years’ reviews:
   1. Are there any difference between two years’ reviews (2017 and 2018)? For example, it can be expectable that one topic was still novel in 2017, but not that novel in 2018. Can you automatically predict 2018’s review opinion given 2017’s reviews?

**== Related Work ==**

[A Dataset of Peer Reviews (PeerRead): Collection, Insights and NLP Applications](https://arxiv.org/abs/1804.09635)

[Open Scholarship and Peer Review: a Time for Experimentation](https://openreview.net/pdf?id=xf0zSBd2iufMg)

Sentiment analysis:

[1] [Mining and Summarizing Customer Reviews](https://www.cs.uic.edu/~liub/publications/kdd04-revSummary.pdf)

[2] [Latent Aspect Rating Analysis on Review Text Data: A Rating Regression Approach](https://www.cs.virginia.edu/~hw5x/paper/rp166f-wang.pdf)

[3] [Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank](http://www.aclweb.org/anthology/D13-1170)