Crawl and Visualize ICLR 2021 OpenReview Data

Descriptions

This Jupyter Notebook contains the data crawled from ICLR 2021 OpenReview webpages and their visualizations. The list of submissions (sorted by the average ratings) can be found here.

Prerequisites

python 3.7

selenium

pandas

seaborn

imageio

wordcloud

tqdm

edgewebdriver NOTE: You can also use chromedriver by setting driver = webdriver.Chrome('chromedriver.exe') .

Crawl Data

Run crawl_paperlist.py to crawl the list of papers (~0.5h). Run crawl_reviews.py to crawl the reviews (~1.5h). NOTE: currently only review ratings are crawled.

Visualization

Keywords Frequency

The top 50 common keywords (uncased) and their frequency:

Keywords Cloud

The word clouds formed by keywords of submissions show the hot topics including deep learning, reinforcement learning, representation learning, graph neural network, etc.

Ratings Distribution

The distribution of reviewer ratings centers around 5 (mean: 5.367).

Keywords vs Ratings

P

The average reviewer ratings and the frequency of keywords indicate that to maximize your chance to get higher ratings would be using the keywords such as deep generative models, or normalizing flows.

All ICLR 2021 Submissions

Number of submissions: 2966 (Collected at 11/11/2020 09:11 AM UTC+8).

Acknowledgment

Visualizations are inspired by this repo: https://github.com/shaohua0116/ICLR2020-OpenReviewData.

Reference

Neuropsychedelia: The Revival of Hallucinogen Research since the Decade of the Brain

Nursing Research: Generating and Assessing Evidence for Nursing Practice