Deep Learning NLP Recommender Challenge

Description

This challenge is to demonstrate the ability to construct a Recommender model using a deep learning architecture using NLP techniques. The dataset consists of (1) user-news interactions and (2) news titles and abstracts. The model should be able to predict whether the user will click on an unseen news article.

Please do not take more than 4 hours of hands-on time for this challenge. The goal is to see how you would approach such a problem and to see how you would code the solution.

Goals

- Design a deep learning network that can take as input user-news items and output a binary variable (1/0) for prediction of user clicking on article
- Build and train the model using a Python deep learning library
- Test the model against user-news combinations not seen in training
- Report on the performance of the model

Instructions

- Read about the dataset at
  https://github.com/msnews/msnews.github.io/blob/master/assets/doc/introduction.md
- Download user-news interactions at
  https://inspire-data-challenge.s3.amazonaws.com/user_news_clicks.csv
- Download news titles and abstracts at
  https://inspire-data-challenge.s3.amazonaws.com/news_text.csv
- NOTE: you do not need to use the entire dataset, if resources are limited. Feel free to sample
- Code a model using a Python library such as FastAI or Pytorch
- Write a brief document describing:
    - Your approach in designing the model
    - Model network architecture

- ○ Training/testing methodology
- ○ Sampling methodology (if applicable)
- ○ Model performance
- ○ Address how you would deploy and monitor the model in production
- ● Also answer the following statistical question (unrelated to the model)
  - ○ You have a client who is testing out a new marketing methodology. In order to test the new methodology, the client splits their users into two groups. In the first group, some users are treated with the new marketing methodology and others are put into a holdout. In the second group, some users are treated with the old marketing methodology and others are put into a holdout. The client sends you a report on the performance of the two campaigns: (their old methodology vs the new methodology) detailing the overall incremental conversion rate above the holdout of each group as well as "statistical significance" of the incremental conversation rate. "The new methodology shows an incremental conversion rate above its holdout of 3.5% with a statistical significance of 99%. The old methodology shows an incremental conversion rate above its holdout of 2.5% with a statistical significance of 99%". The client asks you if we can conclude whether the new marketing methodology will provide an increase in incremental conversions over the old methodology? This is all of the information you have to work with. You will need to interpret and come up with your conclusions from these two statements.