

Yelp Dataset Challenge

Hua Wei
2018-Feb-04

Motivation

- Want to know where to get the best food for you?



- The restaurant with highest predicted rating!
- Predict “stars” before review by using “review” + “user” + “business”

Data

- Main dataset: review.json
- Review score, review content...
- Other datasets:
 1. business.json
 2. user.json
 3. checkin.json
 4. tip.json
 5. photos.json
- Many ideas deserved to be studied!
- Graph analysis for user
- Basic ngram statistics for texts
- Distances analysis for texts (review content, tip content...)
- Models: Traditional (SVM, etc.) vs DeepAI (RNN+1FC) ?
- ...