CPSC 330 Lecture 18: Natural Language Processing

Announcements

- Midterm 2 is next week (Nov 14/15)

- More information on Midterm 2, including a practice midterm will be released later tonight

- Reminder: Please double check your Midterm 2 CBTF booking!

UBC
Computer
Science

Recap: Recommender Systems

How the heck did I split the data last class!?

Beyond Error Rate in Recommender Systems

- If a system gives the best RMSE it doesn't necessarily mean that it's going to give best recommendations.

- In recommendation systems we do not have ground truth in the sense that there is no notion of "perfect" recommendations.

- Training your model and evaluating it offline is not ideal.

UBC
Computer
Science

Beyond Error Rate in Recommender Systems

- Other aspects such as simplicity, interpretation, code maintainability are equally (if not more) important than best validation error.

- Winning system of Netflix Challenge was never adopted.
  - Big mess of ensembles was not really maintainable

- There are also other considerations:
  - diversity
  - freshness
  - trust
  - persistence

iClicker Exercise 17.2

Select all of the following statements which are True (iClicker)

- a. In content-based filtering we leverage available item features in addition to similarity between users.

- b. In content-based filtering you represent each user in terms of known features of items.

- c. In the set up of content-based filtering we discussed, if you have a new movie, you would have problems predicting ratings for that movie.

- d. In content-based filtering if a user has a number of ratings in the training utility matrix but does not have any ratings in the validation utility matrix then we won't be able to calculate RMSE for the validation utility matrix.

Introduction to Natural Language Processing

Natural Language Processing (NLP) is a branch of machine learning focused on enabling computers to understand, interpret, and generate human language.

UBC
Computer
Science

Motivation and Context

- Do large language models, such as ChatGPT, "understand" your questions to some extent and provide useful responses?

- What is required for a machine to "understand" language?

- So far we have been talking about sentence or document representations.

- Today, we'll go one step back and talk about word representations.

UBC
Computer
Science

Referential Ambiguity

Let's start with this picture:

If the **baby** does not thrive on raw **milk**, boil **it**.

it = ?

- How do we know what the it is referring to?

- How do we tell an algorithm that?

Activity: Context and word meaning

Pair up with the person next to you and try to guess the meanings of two made-up words: `flibbertigibbet` and `groak`.

1. The plot twist was totally unexpected, making it a **flibbertigibbet** experience.
2. Despite its **groak** special effects, the storyline captivated my attention till the end.
3. I found the character development rather **groak**, failing to evoke empathy.
4. The cinematography is **flibbertigibbet**, showcasing breathtaking landscapes.
5. A **groak** narrative that could have been saved with better direction.
6. This movie offers a **flibbertigibbet** blend of humour and action, a must-watch.
7. Sadly, the movie's potential was overshadowed by its **groak** pacing.
8. The soundtrack complemented the film's theme perfectly, adding to its **flibbertigibbet** charm.
9. It's rare to see such a **flibbertigibbet** performance by the lead actor.
10. Despite high expectations, the film turned out to be quite **groak**.
11. **Flibbertigibbet** dialogues and a gripping plot make this movie stand out.
12. The film's **groak** screenplay left much to be desired.

UBC
Computer
Science

Demo

You can follow along here!