Weekly Updates on the Recommender System for Toloka

Mitanshi Vyas
Master's Student
04/14/2023

TABLE OF CONTENTS

01
Similarity based Recommender System

02
Sentence Transformer based Recommender System

03
Comparison and Evaluation of both

04
Conclusion

Overview for All The Models

- I have used three different approaches for generating recommendations, including:

    a. Similarity matrices
    b. Classification models
    c. Sentence transformers

- The results and findings from each approach and discuss their strengths and weaknesses.

- Compare the performance of each approach and analyze their respective advantages and limitations.

- In conclusion, I will summarize my key findings and potential areas for further research and development.

Classification Models Approach

- **Recap:**

  I employed four classification models and received best F1 Score of 0.67 from Gaussian Naive Bayes model

Issues:

- Major assumption that all the tasks were displayed to each user

  Why is it problematic?

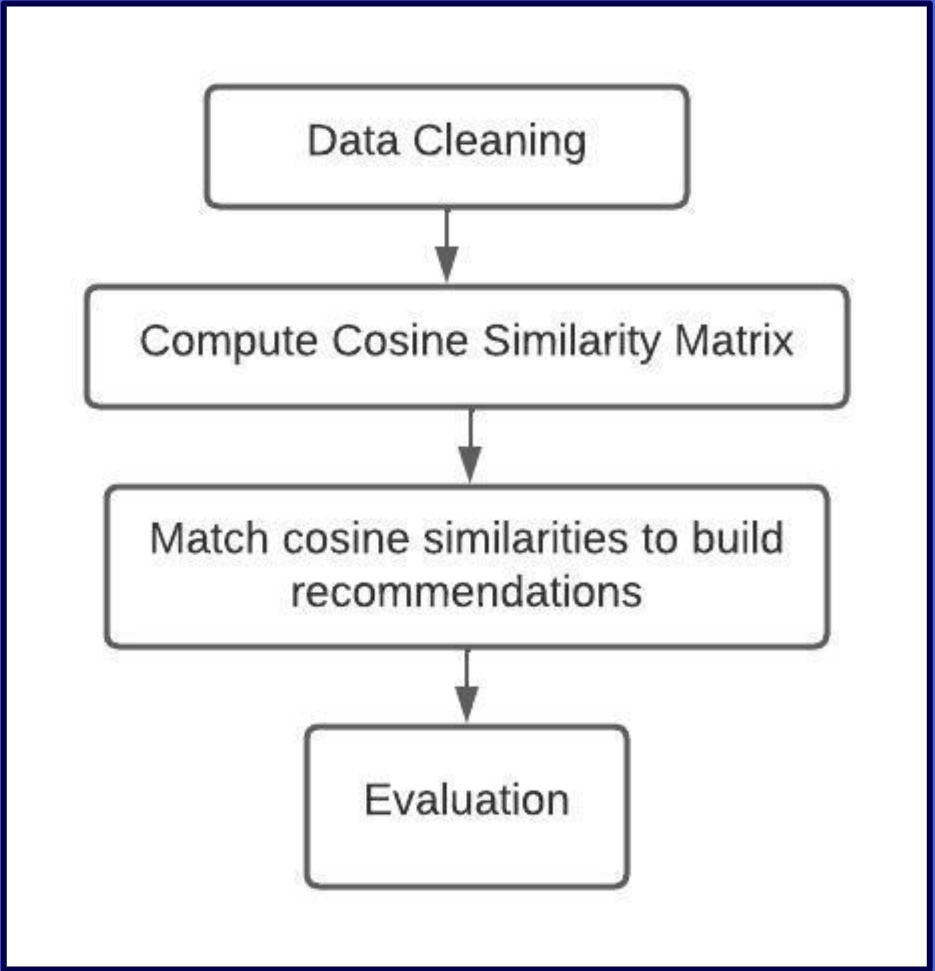
  1. Although we do have the data about which worker finished what task, however, we don't have any information about which tasks were being displayed to each worker
  2. Considering this major assumption can lead to faulty predictions

- We already have very small dataset and since we are splitting the dataset for evaluation purposes, it will further decrease the size of our data

In conclusion, not the best method for building recommendations as compared to others.

Similarity Matrices

I employed cosine similarity matrix, building matrices for all the embeddings or feature vectors. This technique helps to identify similar items or users, which can then be used to make personalized recommendations.



EVALUATION

- For evaluation of models built using similarity matrices, I am using two evaluation metrics:
  a. MAP@k
  b. NDCG@k

These are the evaluation metrics that measure the quality of the recommendations generated by a recommender system.

- MAP@k (Mean Average Precision at K) and NDCG@k (Normalized Discounted Cumulative Gain at K) are used to evaluate the performance of a recommender system, taking into account both the relevance and the order of the recommendations.
- Which means that the system is penalized for recommending irrelevant items higher up the list

RESULTS

```
evaluate_model()

(0.46666666666666673, 0.5156223656766757)
```

A MAP@K value of 0.46 means that, on average, the top K recommendations provided by the system have a 46% chance of being relevant to the user.

An NDCG@K value of 0.51 means that the system's top K recommendations, on average, are about 51% as good as the ideal recommendations that could be provided.

Both of these metrics indicate that the system is performing reasonably well, but there is still room for improvement.

Recommendations

```
User ID:  1316040758
 Porject ID  Title
------------  --------------------------------------
       81077  Product Image Validation V1
       43625  Validation of English Ad Description
       71859  Definitive Image Labeling - EN
       80983  Product Price Validation V1
       74976  MMRepresentativeImageV2
----------------------------------
```

```
----------------------------------
User ID:  1430094402
 Porject ID  Title
------------  ----------------------------------------
      121035  Новостность документа
      102454  Модерация фотографий отельных номеров
       77419  Распознавание таблиц
       23183  Модерация названий организаций
       22597  Модерация описаний исполнителей
----------------------------------
```

Recommendations

```
--------------------------------
User ID:  1490584733
  Porject ID  Title
------------  ----------------------------------------------------------------------------
       71859  Definitive Image Labeling - EN
       74976  MMRepresentativeImageV2
       93220  Does coupon screenshot match the provided coupon's information? (New Zealand)
       81077  Product Image Validation V1
       81980  Image Query Result Relevance T2 MMIR English
--------------------------------
```

```
--------------------------------
User ID:  1699065816
  Porject ID  Title
------------  ----------------------------------------------
       89348  Crowd Spam Labeling - Spanish
       62518  [Toloka] Prueba de comprensión de español
       79091  Crowd Spam Labeling - English
      113583  Comparar resultados de búsqueda – Español - V2
      105766  [Toloka] Prueba de gramática española
--------------------------------
```

Utilizing Sentence Transformers

- We are using Sentence Transformers to convert the text descriptions and titles of the projects into dense vectors (also known as embeddings) in a high-dimensional space.
- These embeddings capture the semantic meaning of the text data and allow us to compute similarities between the projects using cosine similarity.
- Using Sentence Transformers improves the quality of recommendations as compared to using simple bag-of-words or TF-IDF vectorization.

RESULTS

```
[5] evaluate_model()

    (0.5, 0.6684454183016579)
```

These MAP@k and NDCG@k values are slightly better than previous ones.

Recommendations

```
--------------------------------
User ID:  1699065816
  Porject ID  Title
------------  ------------------------------------------------------------
       79091  Crowd Spam Labeling - English
      101570  Search for businesses by specific category (English)
       89348  Crowd Spam Labeling - Spanish
       89350  Crowd Spam Labeling - Russian
       62518  [Toloka] Prueba de comprensión de español
--------------------------------
```

```
--------------------------------
User ID:  1703905621
  Porject ID  Title
------------  ------------------------------------------------------------
       74976  MMRepresentativeImageV2
      113582  Сравните Результаты Поиска – Русский язык - V2-
      110607  Image Relevance Superfresh V2 SFMMIR - English
       72322  MMWatermarkImageDetection
      107829  Video Relevance Labeling (English) - New and improved
--------------------------------
```

Recommendations

```
User ID:  1706440161
 Porject ID  Title
-----------  ------------------------------------------------------
      74976  MMRepresentativeImageV2
      70834  Отбор на задания классификации текста и изображения
      93218  Get coupon information from websites (New Zealand)
     102966  Check the similarity between two products
      78414  Check the similarity between two products
--------------------------------
```

```
--------------------------------
User ID:  1718146027
 Porject ID  Title
-----------  ----------------------------------------------
      74976  MMRepresentativeImageV2
     106175  Does the store have this product? (English)
      72322  MMWatermarkImageDetection
     110607  Image Relevance Superfresh V2 SFMMIR - English
      78414  Check the similarity between two products
--------------------------------
```

Final Thoughts

- Best F1 score from classification model was provided by Gaussian Naive Bayes
- However, Classification model approach isn't the most reliable approach considering the major assumption we had to take
- Received decent values of MAP@k and NDCG@k from ustilizing similarity matrices
- These values further improved with the utilization of Sentence Transformers as they consider the semantic meaning of the text data

THANK YOU!