CS 579 Project 2 - Fairness in Recommender System

Venkata Sai Naveena Bathula(A20451804)

Jesleen Sonia Pradeep Kamalesh(A20448891)

Date: 30 Nov 2020,Group 27

1. Introduction

Fairness in recommendation:

Recommender systems are related to classification and ranking. Fairness is achieved when the recommender compiles a set of objects, such that the ratio of objects from various groups (output) is the same as the ratio present in the subject's input preferences (input). Most recommender system typically learn from past user interactions and preferences to recommend items (movies, products, etc.) to users. The success of a recommender algorithm is generally evaluated with the accuracy of its recommendations, that is, how well the algorithm predicts whether a user will like an item or not- its utility. Fairness in recommendation can be divided into many categories. Three categories: fairness for subjects, fairness for objects, and multi-stakeholder recommendation.

Fairness for subjects:

Considering the group fairness for subjects in memory-based collaborative filtering techniques predict ratings (e.g., matrix factorization). Fairness is achieved at the ideal state in which the protected group incurs rating prediction errors in parity with the non-protected group. Deviation from this ideal state is quantified with four different metrics, which are introduced in the learning process as regularization terms. When aggregating other subjects' opinions, as in user-based collaborative filtering, the idea is to ensure the protected group is well represented.

Fairness for objects:

This may help to picture groups of objects in analogy to item categories, like movie genres. For a particular subject, a list of recommendations is fair (called calibrated) if it contains objects of various groups with a ratio that is equal to the group ratio present in the subject's input preferences. For example, if a user has liked 7 romance and 3 action movies, a recommended list is fair if it contains 70% romance and 30% action movies. Deviation from this state is quantified by a distance metric (KL divergence). To ensure fair recommendations, a greedy iterative re-ranking approach (post-processing), in the spirit of MMR, is taken. The goal is to construct a list that balances the utility of the objects selected and the list's deviation from the input preferences.

Multi-stakeholder recommendation:

A multi-stakeholder recommender system is one in which the objectives of multiple parties, in addition to objectives attributed to the user, are considered in the computation of recommendations. Especially a system in which such parties lie on different sides of the recommendation interaction. For example, at the same time that the system is ensuring that male and female users to get recommendations with similar salary distributions, it might also need to ensure that jobs at minority-owned businesses are being recommended to the most desirable job candidates at the same rate as jobs at white-owned businesses.

Calibrated recommendation:

Calibration is a post-processing technique to improve error distribution of a predictive model. Calibration is a general concept in machine learning, and recently experienced a resurgence in the context of fairness of machine learning algorithms. A classification algorithm is called calibrated if the predicted proportions of the various classes agree with the actual proportions of data points in the available data. The goal of calibrated recommendations is to reflect the various interests of a user in the recommended list, and with their appropriate proportions

For example, when a user has watched, say, 70% romance movies and 30% action movies in the past, then it is reasonable to expect the personalized list of recommended movies to be comprised of 70% romance and 30% action movies as well since we would like to cover the user's diverse set of interests. A recommendation that actually reflect most if not all of the user's interest is considered a Calibrated Recommendation**.**

Goal of the project:

The project goal is to achieve fairness for objects in the recommender system implementing similar methods in calibrated recommendations as post-processing but use different recommender approaches and datasets.

Approaches:

Distributions over genres of use's history and recommended list:

$$p(g|u) = \frac{\sum_{i \in \mathcal{H}} w_{u,i} \cdot p(g|i)}{\sum_{i \in \mathcal{H}} w_{u,i}}, \quad q(g|u) = \frac{\sum_{i \in \mathcal{I}} w_{r(i)} \cdot p(g|i)}{\sum_{i \in \mathcal{I}} w_{r(i)}},$$

To get the similarity between these two distributions, we use Kullback-Leibler divergence (or Jensen-Shannon Divergence):

$$C_{\mathrm{KL}}(p, q) = \mathrm{KL}(p||\tilde{q}) = \sum_g p(g|u) \log \frac{p(g|u)}{\tilde{q}(g|u)},$$

To determine the recommendation list *I*:

$$\mathcal{I}^* = \underset{\mathcal{I}, |\mathcal{I}|=N}{\arg\max}\ (1-\lambda) \cdot s(\mathcal{I}) - \lambda \cdot C_{\mathrm{KL}}(p, q(\mathcal{I}))$$

Description of the dataset:

- We have considered the yelp hotel dataset which contains yelp hotel ratings, reviewer information and hotel information.

Useful attributes from the dataset:

review table- reviewID, reviewerID, hotelID, reviewContent
hotel table- hotelID, name, location, rating, PriceRange, categories.

Usage:

- 'Hotel table' and 'Review table' is considered as the important table for the caliberative filtering of the data.

Brief introduction of related techniques:

- Collaborative filtering is used by recommender systems in which the process of filtering for information or patterns using techniques involving collaboration among multiple agents, viewpoints, data sources, etc. Applications of collaborative filtering typically involve very large data sets.
- Collaborative filtering systems have many forms, but many common systems can be reduced to two steps:
  - Look for users who share the same rating patterns with the active user.
  - Use the ratings from those like-minded users found to calculate a prediction for the active user

- Collaborative filtering is a class of recommenders that leverage only the past user-item interactions in the form of a ratings matrix. It operates under the assumption that similar users will have similar likes.
- 'Categories' is considered as the important attribute for the hotel table which is useful for the calibrated recommendation.
- Calibration is especially important in light of the fact that recommender systems optimized toward accuracy (e.g., ranking metrics) in the usual offline-setting can easily lead to recommendations where the lesser interests of a user get crowded out by the user's main interests.

2. Experiment

Pre-processing of data:

The raw “yelphoteldata.db” contains user's rating for each hotel. Here, we will focus on implicit data, and follow the procedure of simulating binary implicit feedback data. Data preprocessing is a crucial step for any data analysis problem. The model's accuracy depends mostly on the quality of the data. During data preprocessing, we connect to the given database and make the data ready by taking review table and hotel table into dataframes and merge.

Item-item collaborative filtering:

Item-Item collaborative filtering look for items that are similar to the hotels that user has already rated and recommend most similar hotels. This is more stable as compared to the User based collaborative filtering because the average item has a lot more ratings than the average user.

Sparse matrix calculation:

In recommender systems, we typically work with very sparse matrices as the item is very large while a single user typically interacts with a very small subset of the item. A sparse matrix is calculated using the important columns ‘review ID’, ‘rating’ and ‘hotelID’ from the yelp hotel data set which is used for ranking using collaborative filtering

Ranking:

The recommender systems trained toward accuracy with ranking metrics can easily generate lists of recommended items that focus on the main areas of interest of a user–while the user’s lesser areas of interest tend to be underrepresented or even absent. Ranking the predicting movies is done using “BayesianPersonalizedRanking’s” recommend method.

KL divergence:

To measure the difference between two probability distributions over the same variable x, a measure, called the Kullback-Leibler divergence, or simply, the KL divergence, has been popularly used in the data mining literature. The concept was originated in probability theory and information theory.

Generating Calibrated Recommendations:

Being able to compute the calibration metric between p(g|u) and q(g|u) is all well and good, but how can we generate a recommendation list that is more calibrated becomes the next important and interesting question.

Different recommendation algorithm's objective function might be completely different, thus instead of going to hard-route of incorporating it into the objective function right off the bat and spend two weeks writing the customized algorithm in an efficient manner, we will start with an alternative approach of re-ranking the predicted list of a recommender system in a post-processing step.

To determine the optimal set I∗ of N recommended items, we'll be using maximum marginal relevance.

$$I^{*} = \operatorname{argmax}_{I,\ |I|=N} (1-\lambda)\cdot s(I) - \lambda\cdot C\,(p, q(I))$$

Where

- s(i) is the score of the items i∈I predicted by the recommender system and s(I)=∑i∈Is(i), i.e. the sum of all the items' score in the recommendation list.
- λ∈[0,1]λ∈[0,1] is a tuning parameter that determines the trade-off between the score generated by the recommender and the calibration score, notice that since the calibration score is measured by KL-divergence, which is a metric that's the lower the better we use its negative in the maximization formula.
- We have considered Lamba value=0.95 and alpha value = 0.03

Re-ranking by calibrated recommendations:

Re-ranking model can be easily deployed as a follow-up modular after any ranking algorithm, by directly using the existing ranking feature vectors. It directly optimizes the whole recommendation list by employing a transformer structure to efficiently encode the information of all items in the list. By implementing re-ranking approach, the calibrated recommendation’s KL divergence is lower than the original recommendation value.

Hyper- parameter tuning:

Before tuning the parameters, we need to pick up an evaluation metric considering Lamba value=0.95 and alpha value = 0.03. A popular evaluation metric for recommenders is Precision at K which looks at the top k recommendations and calculates what proportion of those recommendations were actually relevant to a user. To find the parameters that give the best precision at K or any other evaluation metric that one wants to optimize. The KL divergence value is 0.06343693676317233.

```
0.06343693676317233
```

Problems encountered:

- Problem- Faced an issue while extracting the data without the table name.
  Accomplishment- We figured out to list the table names by selecting from sqlite_master

- Problem- There was a problem finding on what basis ranking the hotels to users based on similarity scores.
  Accomplishment- Tried few solutions but figured out cosine similarity will produce a better result.

- Problem –Memory Error was occurring during the creation of the item matrix for collaborative filtering due to the amount of data in the dataset.
  Accomplishment – By using the sparse matrix, memory error was avoided.

3. Experiment Results

- After performing pre-procesing, the useful attributes are obtained as shown

	reviewerID	hotelID	rating	categories
0	IFTr6_6NI4CgCVavIL9k5g	tQfLGoolUMu2J0igcWcoZg	5	Event Planning & Services, Hotels, Hotels & Tr...
1	c_-hF15XgNhlyy_TqzmdaA	tQfLGoolUMu2J0igcWcoZg	3	Event Planning & Services, Hotels, Hotels & Tr...
2	CiwZ6S5ZizAFL5gypf8tLA	tQfLGoolUMu2J0igcWcoZg	5	Event Planning & Services, Hotels, Hotels & Tr...
3	nf3q2h-kSQoZK2jBY92FOg	tQfLGoolUMu2J0igcWcoZg	1	Event Planning & Services, Hotels, Hotels & Tr...
4	Sb3DJGdZ4Rq__CqxPbae-g	tQfLGoolUMu2J0igcWcoZg	3	Event Planning & Services, Hotels, Hotels & Tr...
...	...	...	...	...
688324	e7B7IsZIRT8LbFj8FcY78w	9xny0IlJqTlnobC6W-UxbA	5	Event Planning & Services, Hotels, Hotels & Tr...
688325	e7B7IsZIRT8LbFj8FcY78w	PmmTXis1gCL34mg2bZ9gtw	2	Event Planning & Services, Hotels, Hotels & Tr...
688326	e7B7IsZIRT8LbFj8FcY78w	Mr6zu_hWk2CodBdqqMWQjg	5	Nightlife, Bars, Dance Clubs, METADATA
688327	e7B7IsZIRT8LbFj8FcY78w	-zetzVfO4X0dpiiTmjdeKg	5	Restaurants, Sushi Bars, Japanese, METADATA
688328	e7B7IsZIRT8LbFj8FcY78w	kEfMOwJRw2CNgPTVh82JaA	5	Beauty and Spas, Hair Salons, METADATA

688096 rows × 4 columns

- The top 10 recommendations using collaborative filtering are

```
Enter top number of recommendations 10

[uqvNkdNnDxGWmgXDV3wKQA,
 3RnY8fcjFboByx4BegpVrw,
 9HUBgeAcE4WH5t6uJo9xaw,
 U_EM2fxxYu6sfWzpkLy7WQ,
 v42UtPH-Hbz1vyWwUzvvUQ,
 Qm_JNFCWd56sVXhLQEx8pA,
 iS11Xr1dny7KFCPgP3hYRg,
 KjAM6Djv_Ac10G2n5b9LmQ,
 XoneurHyGGK_KvidCn9tNA,
 QpigCjLGYM_lskKqjHa1cw]
```

- The KL divergence value is noted as

```
calc_KL(visited_hotels, recommended_hotels, 0.03)

2.1433946259825585
```

- The KL divergences values obtained after the hyperparamater tuning are

	lambda	alpha	kl_divergence	calibrated_recommendation
0	0.15	0.01	3.164266	[0-VPL4kl699tEzipR2nteQ, iFVOfHelzTRNVX5xgL_9E...
1	0.15	0.06	1.920729	[0-VPL4kl699tEzipR2nteQ, iFVOfHelzTRNVX5xgL_9E...
2	0.16	0.01	3.164266	[0-VPL4kl699tEzipR2nteQ, iFVOfHelzTRNVX5xgL_9E...
3	0.16	0.06	1.920729	[0-VPL4kl699tEzipR2nteQ, iFVOfHelzTRNVX5xgL_9E...
4	0.17	0.01	3.164266	[0-VPL4kl699tEzipR2nteQ, iFVOfHelzTRNVX5xgL_9E...
...	...	...	...	...
161	0.95	0.06	1.920729	[0-VPL4kl699tEzipR2nteQ, iFVOfHelzTRNVX5xgL_9E...
162	0.96	0.01	3.164266	[0-VPL4kl699tEzipR2nteQ, iFVOfHelzTRNVX5xgL_9E...
163	0.96	0.06	1.920729	[0-VPL4kl699tEzipR2nteQ, iFVOfHelzTRNVX5xgL_9E...
164	0.97	0.01	3.164266	[0-VPL4kl699tEzipR2nteQ, iFVOfHelzTRNVX5xgL_9E...
165	0.97	0.06	1.920729	[0-VPL4kl699tEzipR2nteQ, iFVOfHelzTRNVX5xgL_9E...

166 rows × 4 columns

- The top 10 Calibrated recommendations are

```
[U_EM2fxxYu6sfWzpkLy7WQ,
 KjAM6Djv_Ac10G2n5b9LmQ,
 QpigCjLGYM_lskKqjHa1cw,
 uqvNkdNnDxGWmgXDV3wKQA,
 9HUBgeAcE4WH5t6uJo9xaw,
 3RnY8fcjFboByx4BegpVrw,
 Qm_JNFCWd56sVXhLQEx8pA,
 XoneurHyGGK_KvidCn9tNA,
 v42UtPH-Hbz1vyWwUzvvUQ,
 iS11Xr1dny7KFCPgP3hYRg]
```

4. Summary and Conclusion

Collaborative filtering relies on the user-item interaction and relies on the concept that similar users like similar things**.** To calculate similarity between two items, we look into the set of items the target user has rated and computes how similar they are to the target item i and then selects k most similar items. By implementing the calibrated recommendation as post-processing, the fairness for the objects in recommender systems is achieved where the the calibrated recommendation's KL divergence is lower than the original recommendation value.

5.References

- http://liu.cs.uic.edu/download/yelp_filter/password:y4Filter
- https://dl.acm.org/doi/pdf/10.1145/3240323.3240372
- https://medium.com/@tomar.ankur287/item-item-collaborative-filtering-recommender-system-in-python-cf3c945fae1e
- https://towardsdatascience.com/why-we-use-sparse-matrices-for-recommender-systems-2ccc9ab698a4