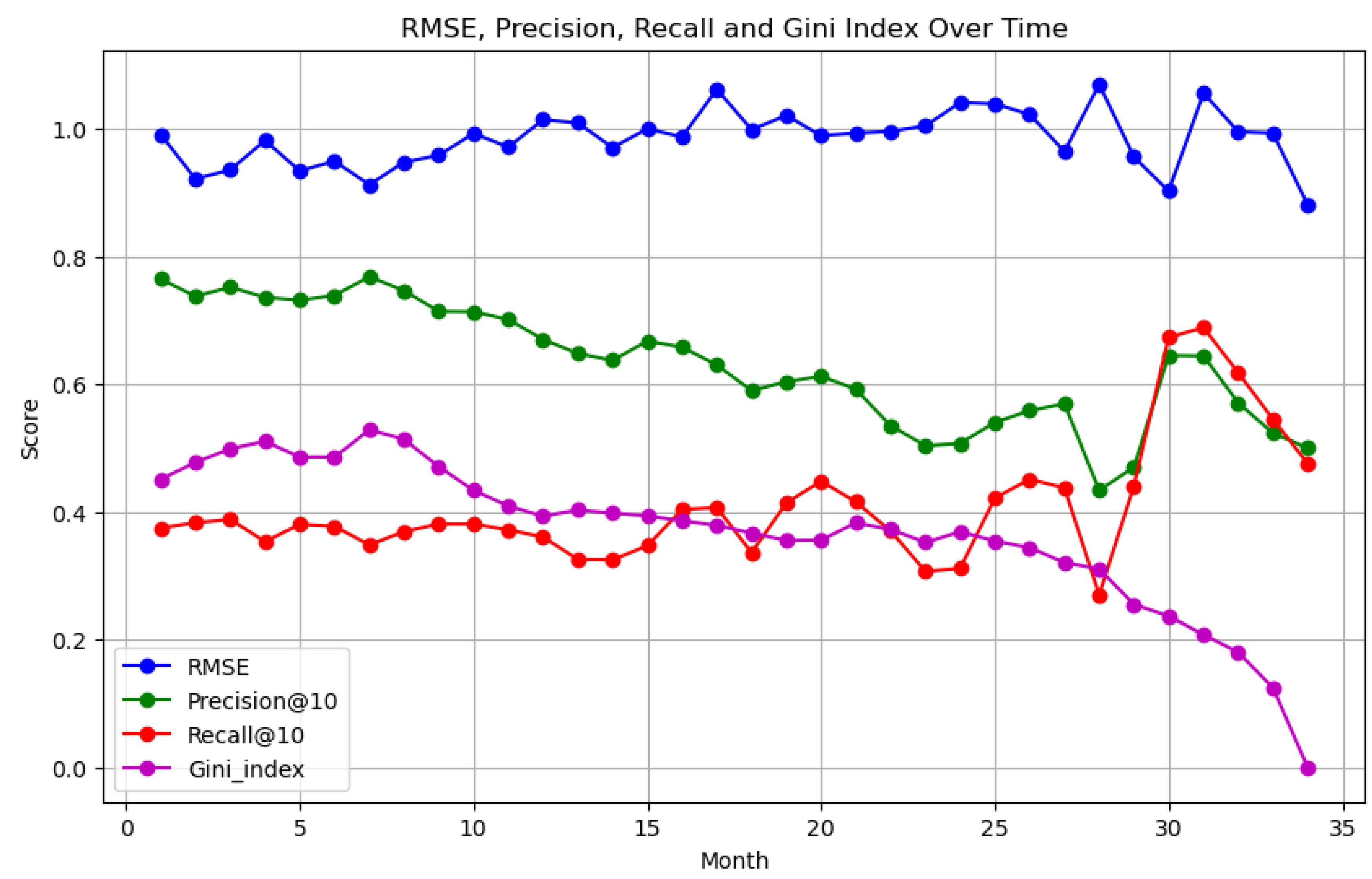
Analysis of Popularity Bias Effect in Media Recommendation

A Longitudinal Study of Popularity Bias in Media Recommendation Systems

Bjørn Kjartansson Mørch, Master's in Information Science, University of Bergen
bmo025@uib.no

Supervisor: Mehdi Elahi, University of Bergen
Co-supervisor: Anastasiia Klimashevskaia, University of Bergen

Media Futures



This graph shows the RMSE, Precision, Recall and Gini Index score over 34 months, which is the full length of the dataset. The colours in the graph are corresponding to the metric shown in the bottom left corner of the figure.

Abstract

Recommender systems are useful tools in media, to help users find what might be interesting to them, and what they might be interested in but not aware of. In real scenarios, recommender systems may focus on recommending only already popular items, with no room for new niche items.

This master's thesis will address this challenge by analysing the longitudinal effect of popularity bias, evaluating how the bias towards popularity differs over a long period over time. To mitigate this popularity bias, the Calibrated Popularity (CP) strategy is applied to provide the individual user with what they want, either popular or niche items, or both.

To evaluate the performances of the base and re-ranked model, different accuracy metrices are used, both normal and beyond. The beyond accuracy metric that will be in most focus of this thesis is the Gini Index, which results in how equal the model is.

Research questions

1. What is the longitudinal effect of popularity bias in media recommendation systems?

2. What is the impact of mitigation strategies in recommendation systems in the long run?

Method

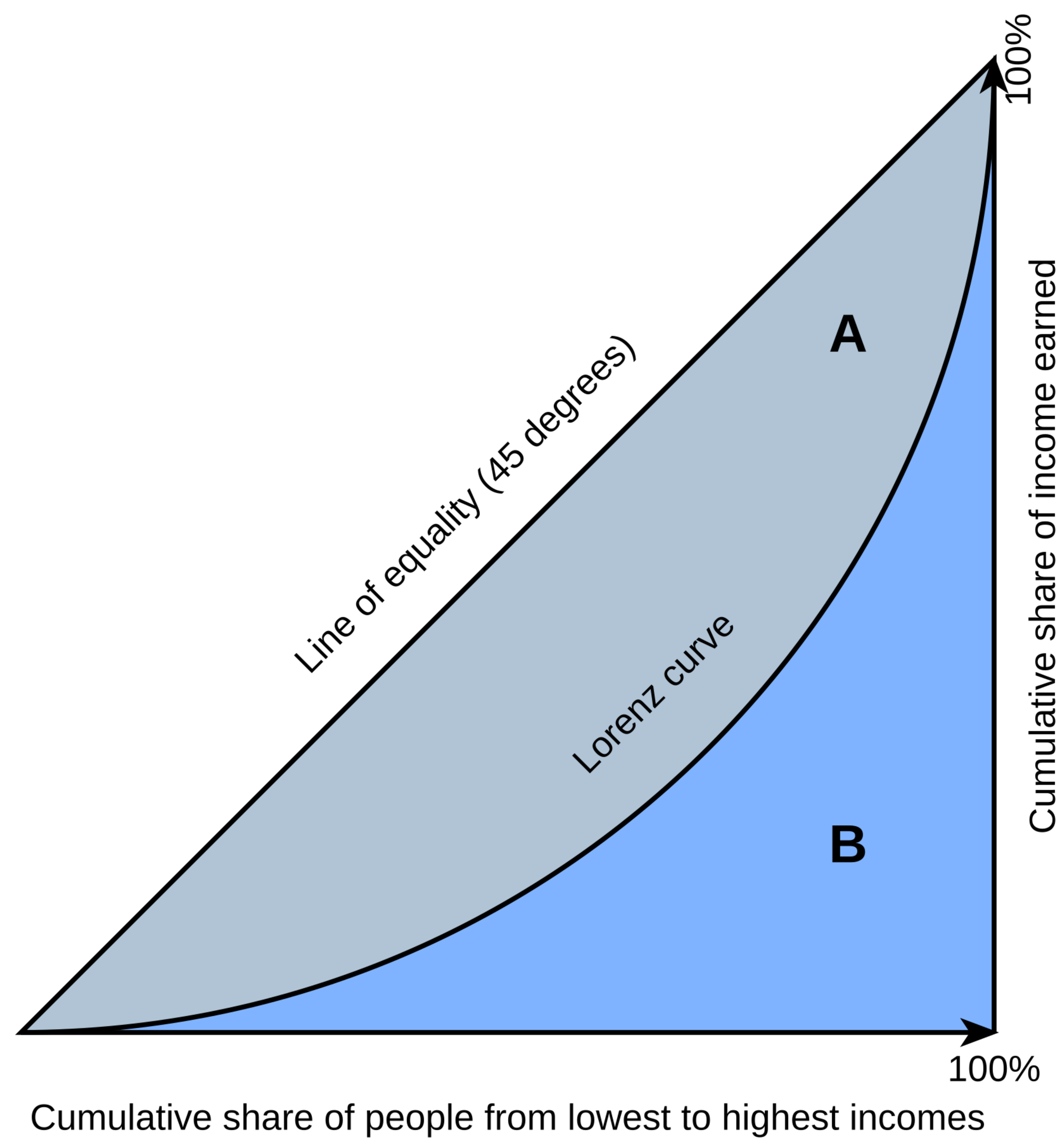
I have used a well-known dataset called MovieLens 1m for experiments of my thesis. The dataset contains user ratings of movies from the period 2000-2003.

In order to evaluate the model dynamically, I sorted the ratings in the dataset according to their timestamp and iterated the evaluation on a monthly basis. The first iteration is the first month as the training set and the second month as the testing set. This continues throughout the entire dataset where the training set is n, and the testing set is n+1, where n is the number of months.

The dataset is firstly evaluated using SVD++ as the base model. Both accuracy metrics, such as RMSE, Precision and Recall, and beyond accuracy metrics, such as the Gini index, Average Recommendation Popularity (ARP) Average Percentage of Long Tail Items (APLT) and Average Coverage of Long Tail items (ACLT), are used.

To mitigate the popularity bias, we use the Calibrated Popularity re-ranking technique. With this technique added to the base model, the user is put into one of three categories; heavily likes popular items, likes both popular items and niche items, and heavily likes niche items. The user is then recommended more of the items that fit their category.

The ultimate goal of the experiments is to find out whether the CP strategy can mitigate the popularity bias effectively by the results of the beyond accuracy metrices.



A simple example of the Gini index; the ratio of area (A) between a perfect equal line and the Lorenz curve.

PARTNERS

amedia Schibsted 2
NRK Bergens Tidende
nb.no vizrt WOLFTECH
NORCE Faktisk.

HOST

UNIVERSITY OF BERGEN

FUNDED BY

This research is funded by SFI MediaFutures partners and the Research Council of Norway (grant number 309339).

Forskningsrådet Norwegian Centre for Research-based Innovation