# Calibrated Recommendations

BY NETFLIX - examples on the topic of movies.

Paper but also video presentation that I will be using.

## Motivation

Based on segmentation of previously played movies, we want to reflect the ratios of genres in the recommended set. That means have all interests of users as well as correct proportions =>> be fair! (apparently a big move towards fairness in the machine learning space in general.)

## Accuracy

THE BE ALL END ALL metric until now - which amplifies main interests

### Example 1 - binary classification

Set of a 100 movies, either romance or action, best accuracy is obtained by assigning all the movies the majority label and not “guessing”.

#### Data set example (paper)

### Example 2 - LDA - Latent Dirichlet allocation

- statistical model
- from Machine learning:
  - a mechanism used for topic extraction.
  - treats documents as probabilistic distribution sets of words or topics.
  - these topics are not strongly defined – as they are identified on the basis of the likelihood of co-occurrences of words contained in them.
  - topic ~ genre, word ~ movie/video

## Calibration metric

- genre distribution - aggregation with weight (decay)
- genre distribution of recommended list - weight based on the rank of the movie in the list
- Using KL divergence (other divergences - Hellinger) “how similar are p and q”

prior - cold start

smoothing term - for q = 0

# Calibration method

Smaller calibration = better score >>> the list we got has the correct proportions of the genres.

This is added to the post processing step as calibration is a list based property, recsys is trained for pointwise/pairwise approach

We run our RecSys, get some score, we subtract by how much it is different from the reference user.

Lambda determines the weight of the calibration in turn determines the tradeoff on the accuracy (remember begining)

PAPER APENDIX - change of terms

In mathematics, a submodular set function (also known as a submodular function) is a set function whose value, informally, has the property that the difference in the incremental value of the function that a single element makes when added to an input set decreases as the size of the input set increases.

Submodular functions have a natural diminishing returns property which makes them suitable for many applications,

# Related concepts

## Fairness

Be fair to genres - no overwhelming of the best for the accuracy

## Diversity

Of the results based on what the user already watched

# Experiments

It improves based on the lambda (wide distribution of Ckl)

## Tradeofs

recall vs calibration (recall - how many relevant items are selected)

# Sumarry

Unbalanced data is normal when training for accuracy
Good for limited data per user

Solution - submodular function in post processing to solve it

# Funny

“I cant talk about AB test result, I cant talk about productionalized things.”

# QA

Interesting question - action VS sport (small number of sport movies) - it would show up sooner + tradeoff.

Set of weights - uniform in paper, but based on the domain.

Fixing what is broken - not done yet

Scaling - on sparse categories better (number of cats, number of items, number of items in the list)

Calibration weight is global for now - can it be personalized?