

Music Recommendation with Collaborative Filtering on Last.fm Data



Karel Beckeringh, s2600358
Wessel Reijngoud, s2580748
Nik van 't Slot, s2536420

Contents

- Introduction
- Materials
- Methodology
- Evaluation
- Results
- Analysis
- Description of individual contributions

Introduction

- Last.fm, website counts listens
- Dataset is rather unique and appeals to us on a personal level
- Friend relations
- User-based, item-based or a combination of techniques
- Earlier research
- What (combination of) recommender technique(s) generates the best results in recommending artists to users?

Materials

- The dataset used is mostly the same as found in HetRec (1)

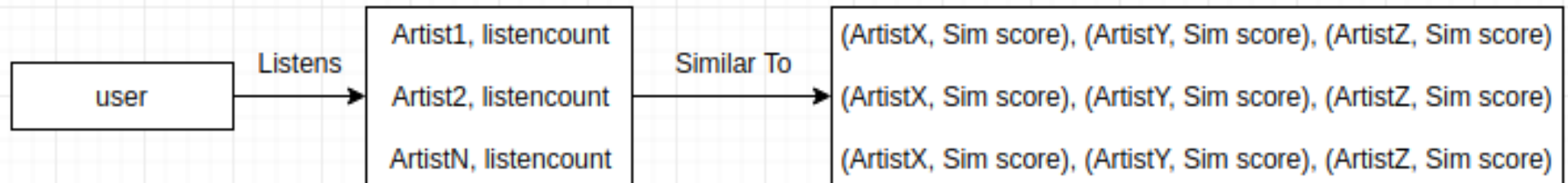
- artists.dat - This file contains information about music artists listened and tagged by the users.
- tags - This file contains the set of tags available in the dataset.
- user_artists.dat - This file contains the artists listened by each user. It also provides a listening count for each [user, artist] pair.
- user_friends.dat - These files contain the friend relations between users in the database.
- test.dat - This file contains 10 %
- training.dat - This file contains the other 90%

Also pickle files created by ourselves that use files shown above to make the program more efficient have been used.

- item_sim50.pickle - This file contains the cosine similarity between artists, only containing those having a score of 0.50 or higher.
- user_sim01.pickle - This file contains the cosine similarity between users, only containing those having a score of 0.01 or higher.

Methodology

Item based



Sorted by listencount

Sorted by similarity score

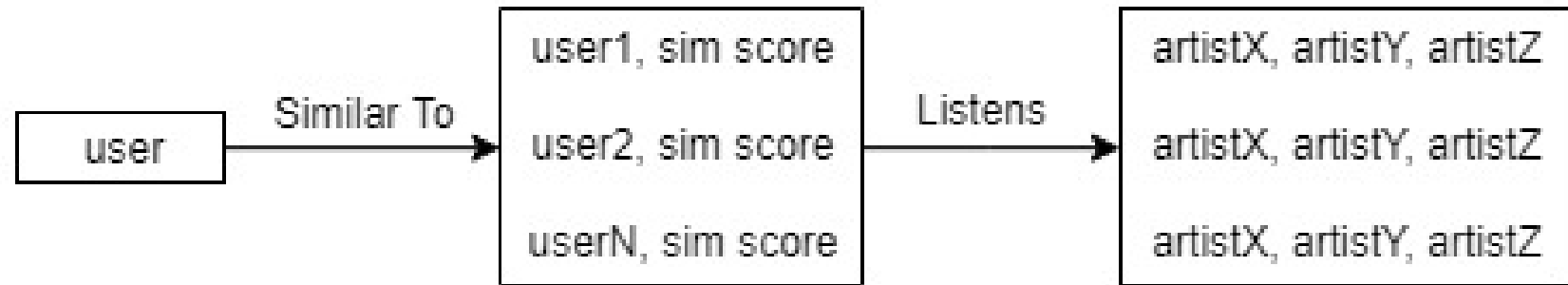
Weight = Listencount * Similarity score

Heighest weights recommended

N=30 most efficient

Methodology

User-based



Sorted by similarity score

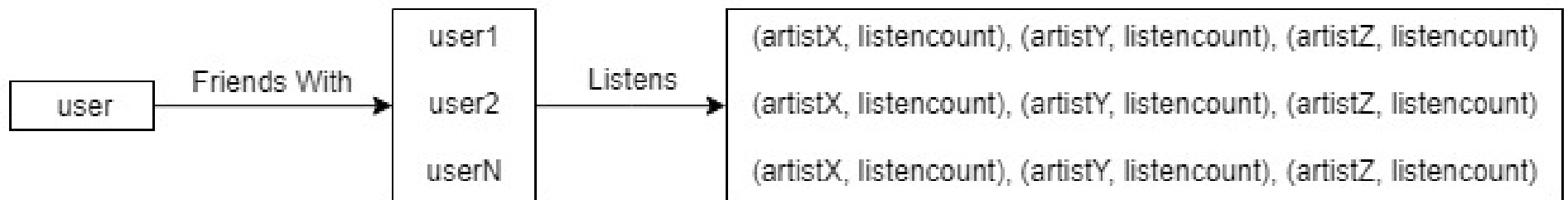
Weight = similarity score

Highest weights recommended

N=34 most accurate

Methodology

Friend-based



Considering all friends, no sorting

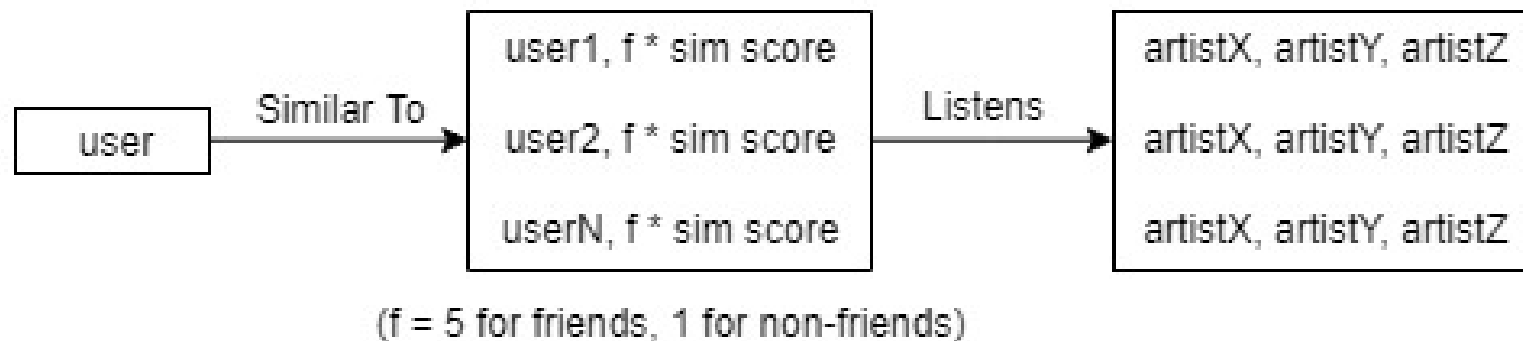
Weight = listencount

Highest weights recommended

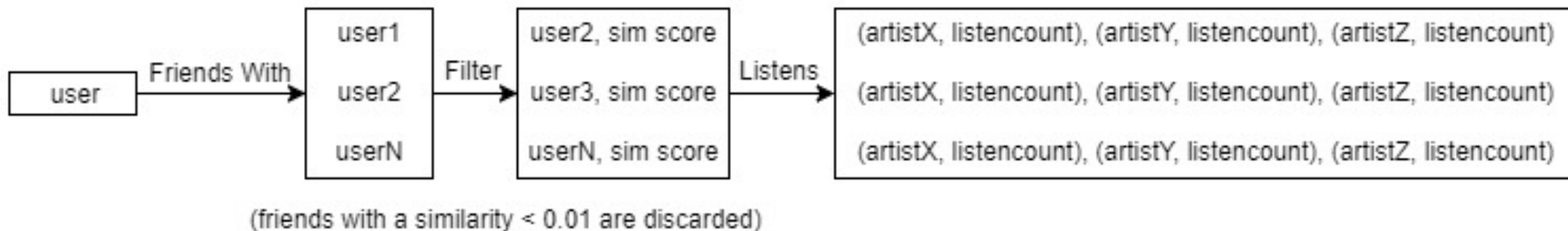
Methodology

Combinations

User-based with friend priority



Friend-based with similarity filter



Evaluation

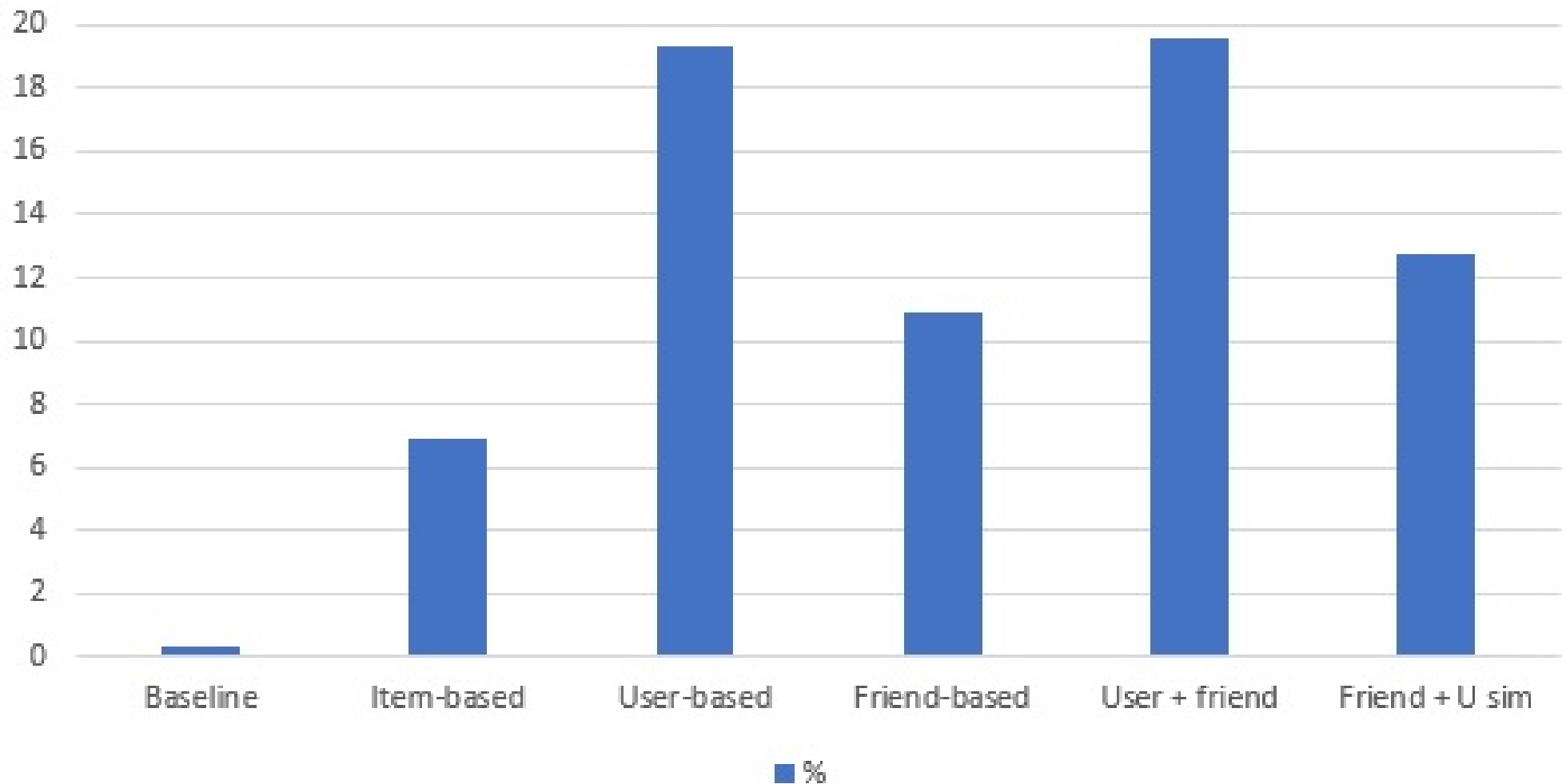
- user_artist.dat split → 10% test 90% training
- Baseline Random recommendation
- Like assignment 3: predict missing artists
- Accuracy = # hits / (# hits + misses)

Results

	Baseline	Item-Based	User-Based	Friend-Based	User + friend	Friend + User
Correct out of 9283	28	643	1794	1010	1823	1186
Accuracy	0.30%	6.93%	19.33%	10.88%	19.64%	12.78%
Time in s	13.127	54.887	20.639	6.929	22.813	6.953

Results

Recommender Scores



Analysis

- All better than baseline
- Item-based low accuracy
- User-based (with friend priority) best results
- Way of evaluating could be better
- False misses

Contributions

- Research proposal used – Karel
- User / Friend based – Karel
- Item based – Karel, Wessel, Nik
- Report – Karel, Wessel, Nik
- Presentation – Karel, Wessel, Nik