ELECTRICAL & COMPUTER ENGINEERING

TEXAS A&M UNIVERSITY

ECEN 765: Machine learning with Networks

Recommendation System based on Collaborative Filtering with focus on Cold Start

Presented By :
Ankit Yadav

Contents

- Introduction
- Objectives
- Recommendation System Overview
- Related research work
- Procedure
  - System Overview
  - Methodology
- Result & Conclusion
- Future Scope

Introduction

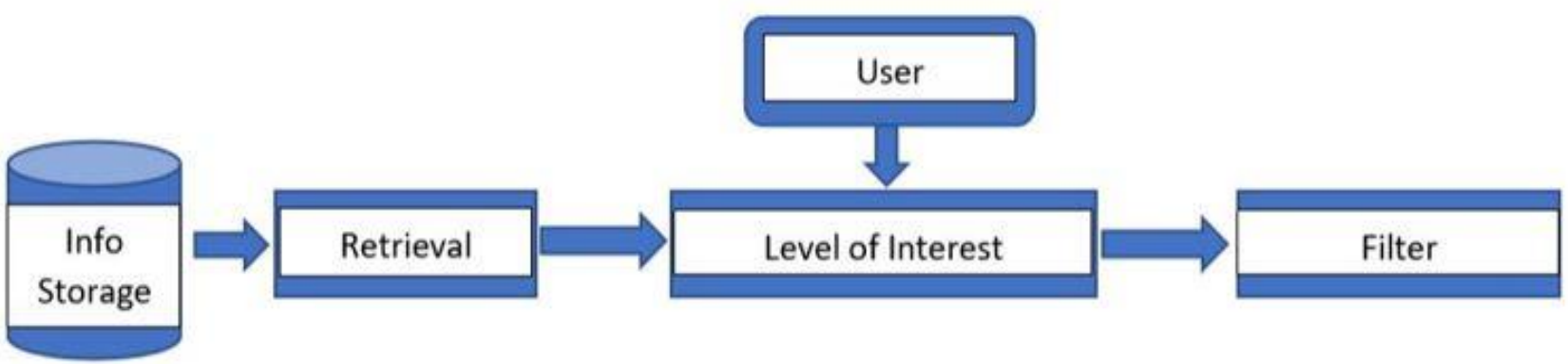
- Internet has witnessed rapid growth
- Addition of new products
- Disinterest of users
- Cold Start Problem

Objective

- Address the cold start problem

Recommendation System

- Recommends items to the user
- First recommendation system built in 1992
- Based on information from the user

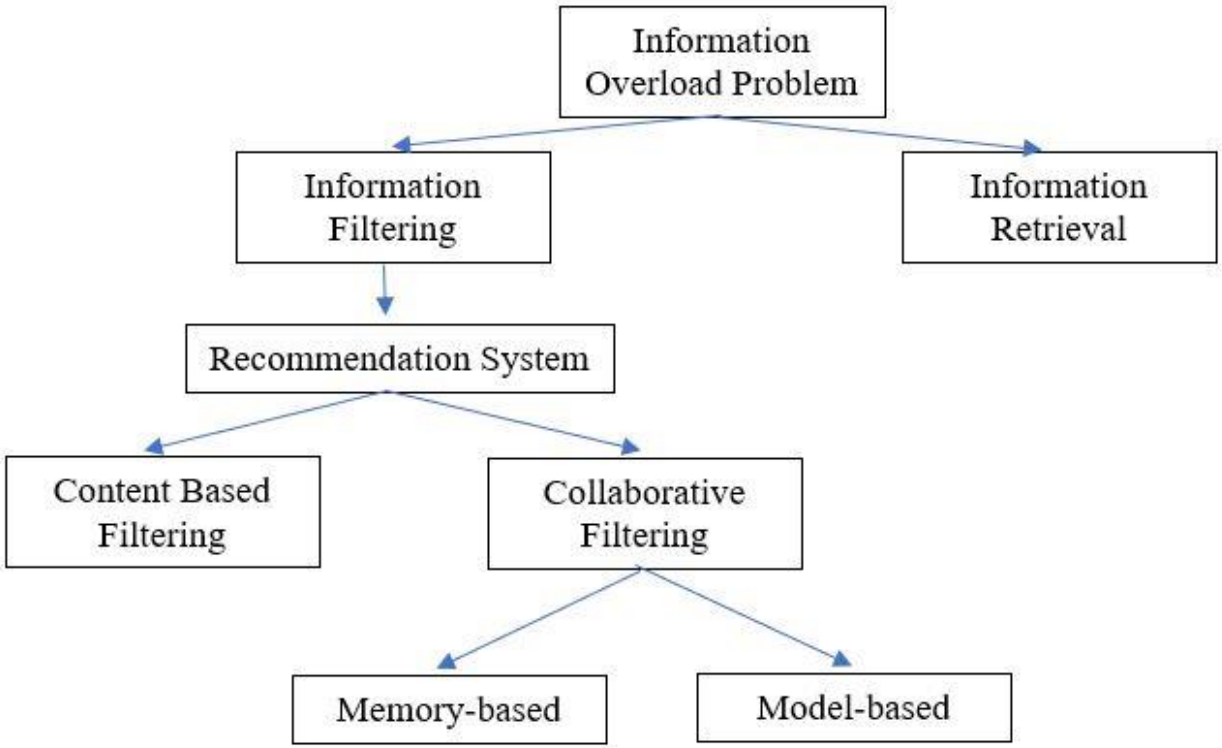


Recommendation System Contd..

Types of Recommendation System:

1. Collaborative Filtering
2. Content based Filtering
3. Demographic Filtering
4. Social Filtering
5. Hybrid Filtering

Information Overload Problem



Related research work

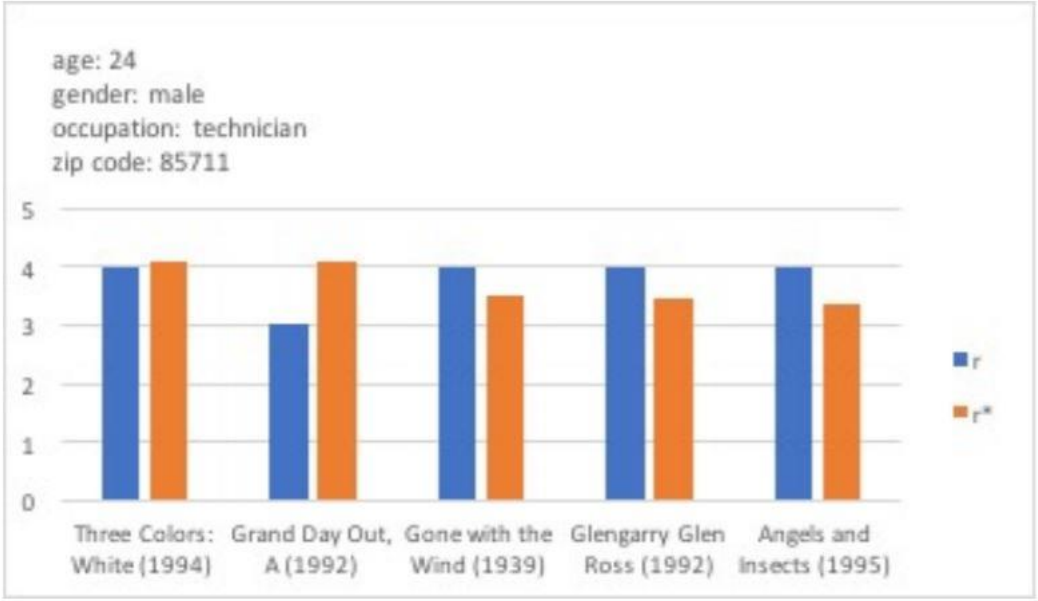
- **Using Demographic Information to Reduce the New User Problem in Recommender Systems**

Demographic Information:

user id | age | gender | occupation | zipcode

User ratings :

user id | item id | rating | timestamp



Related Research Contd..

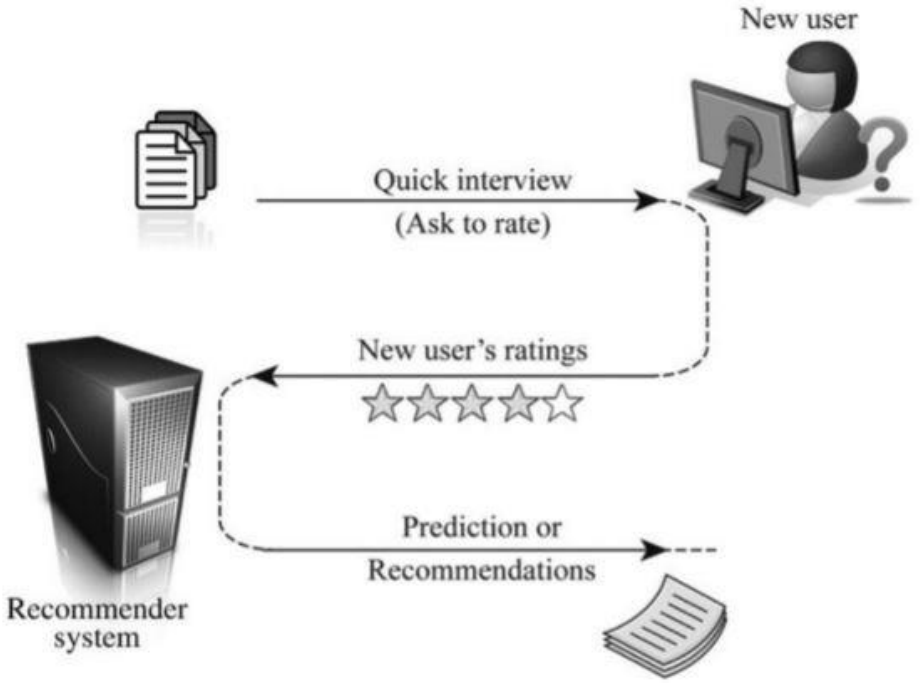
Collaborative Filtering Enhanced By Demographic Correlation

- Introduced two algorithms, U-Demo and I-Demo
- Improves performance

feature #	feature contents	comments
1	age <= 18	• each user belongs to a single age group, • the corresponding slot takes value 1 (true) • the rest of the features remain 0 (false)
2	18 < age <= 29	
3	29 < age <= 49	
4	age > 49	
5	male	• the slot describing the user gender is 1 • the other slot takes a value of 0
6	female	
7-27	occupation	• a single slot describing the user occupation is 1 • the rest of the slots remain 0

Related Research Contd..

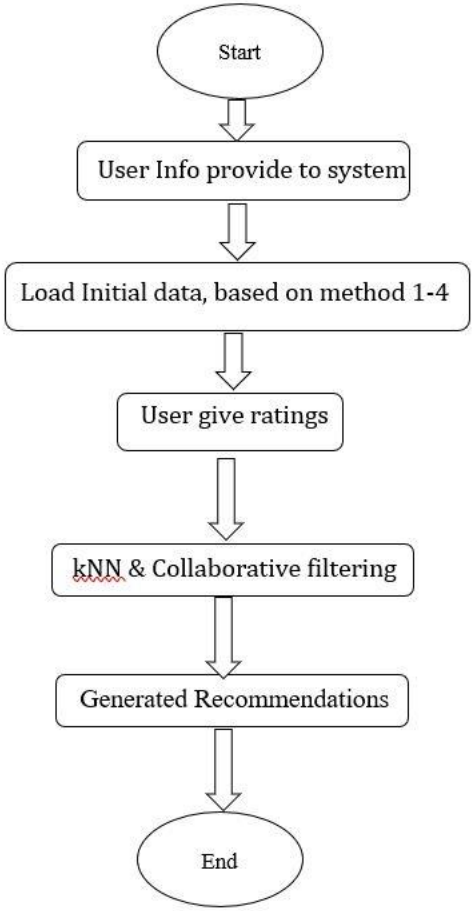
Cold-start Problem in Collaborative Recommender Systems: Efficient Methods Based on Ask-to-rate Technique



Procedure : System overview

- Combination of the above methods
- **Dataset:** MovieLens 100k

- User id | age | gender | occupation | Zipcode
- User id | movie id | rating value



Procedure : System overview

- Preprocessing of the data

1. Age has been divided into the following group: 0-18, 19-24, 25-30, 31-40, 41-50, 51-60, 61-70, 71-100
2. 0/1 corresponding to M/F for gender identification.
3. Occupation has 21 categories in the dataset.

Processed data: **[Age | Gender | Occupation]**

Example:

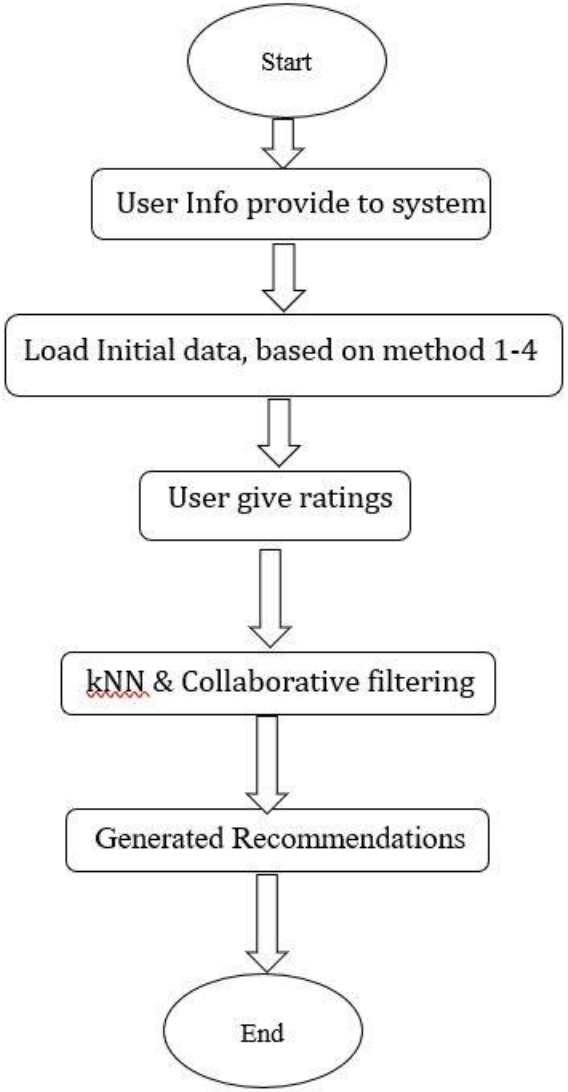
1. **10 year guy**

[1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0]

2. **23 year old, female scientist**

[0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0]

Procedure : Methodology



Result & Conclusion

- Tested it on 20 people.
- Type1: Demography based
- Type2: Entropy0
- Type3: Hybrid
- Type4: Random

User Demographics			User Preference Score			
Age	Gender	Occupation	Type 1	Type 2	Type 3	Type 4
30	M	Engineer	3	4	1	2
22	F	Student	1	3	4	2
38	M	Technician	4	3	2	1
18	M	Student	4	3	1	2
20	M	Artist	4	3	2	1
24	M	Engineer (Just Graduated)	3	4	1	2
43	F	homemaker	3	4	2	1
39	F	homemaker	1	4	2	3
35	M	Other (Gardner)	4	1	2	3
23	F	Student	3	4	1	2
65	M	retired	4	1	2	3

Result & Conclusion

- Entropy0 and demography based system performed almost similar.
- Hybrid system was not as good as the Entropy0 and demography

User Demographics			User Preference Score			
Age	Gender	Occupation	Type 1	Type 2	Type 3	Type 4
30	M	Engineer	3	4	1	2
22	F	Student	1	3	4	2
38	M	Technician	4	3	2	1
18	M	Student	4	3	1	2
20	M	Artist	4	3	2	1
24	M	Engineer (Just Graduated)	3	4	1	2
43	F	homemaker	3	4	2	1
39	F	homemaker	1	4	2	3
35	M	Other (Gardner)	4	1	2	3
23	F	Student	3	4	1	2
65	M	retired	4	1	2	3

Future Work

- Find a new dataset or develop one with more demographic information e.g. religion, ethnicity
- Try the algorithm on other products, not just movies

ELECTRICAL & COMPUTER ENGINEERING

TEXAS A&M UNIVERSITY