Product Recommender System

CS-25
Under the guidance of
Mr. Rajesh Tripathi

Computer Science and Engineering Department
Motilal Nehru National Institute of Technology,Allahabad
1 May 2018

Presented By

Name	Registration No.
Sajad Ansari	20154175
Rajeev Ranjan	20154167
Naval Kishore	20154171
Vaibhav Badole	20154033
Rebba Prashanth	20154150

Overview

- Introduction
- Dataset and preprocessing
- Implementation
- Result Analysis
- Conclusion and Future Work
- References

Introduction

- A Recommender system is a system performing information filtering to bring information items such as movies, music, books, news, images, web pages, tools for the interest of a user.
- The aim of a recommender system is often to help consumers learn about new products and desirable ones among myriad of choices.
- We are interacting with the recommender systems in our day-to-day life like product recommendation in e-commerce sites (Amazon, Flipkart), friend recommendation in social networking sites (Facebook, Instagram) etc.

Dataset and preprocessing

- The dataset contains product reviews and metadata from Amazon, including 142.8 million reviews spanning May 1996 - July 2014.
- There are various categories in Amazon dataset like Books, Electronics, Home-products, Kindle, Cell-phones, Sports and Outdoors, Health and Personal cares etc.[5]
- Each product has more than 10 ratings from different users.The ratings for each product are from 1 to 5.

Dataset and Preprocessing

Sample Review of Dataset :

reviewerID	A2SUAM1J3GNN3B
asin	0000013714
reviewerName	J. McDonald
helpful	[2, 3]
reviewText	I bought this for my husband ....!
summary	Heavenly Highway Hymns
unixReviewTime	1252800000
reviewTime	09 13, 2009

Dataset and Preprocessing

- Creation of dataframe from json file of dataset and conversion of json file into csv file for future processing.
- Listing total unique number of users and products.
- Counting number of reviews per product.
- Extraction of helpful-numerator and helpful-denominator from helpful column.
- Description of mathematical features like mean ,count, standard deviation etc. of dataset to get some insights.

Implementation

Implementation of different types of recommender systems in python using collaborative filtering[6] as learning model.
The different types of recommender systems are implemented as follows.

- Item-Item Based Collaborative Filtering
- User-User Based Collaborative Filtering
- Popularity Model

Implementation

Item-Item Based Collaborative Filtering[2]:

- It is a form of collaborative filtering for recommender systems based on the similarity between items calculated using people's ratings of those items.
- Calculation of similarity between items based on their **Summary Review Text.**
- When users consume and then rate an item, that item's similar items are picked from the existing system model and added to the user's recommendations.

Implementation

Algorithm Proposed: K nearest neighbours classifier

- K nearest neighbors is an algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g. distance functions).
- The nearest neighbors model computes the distance between every interaction vector in the query set against every interaction vector in the reference set. For each vector in the query set, the k closest reference vectors are returned.

Implementation

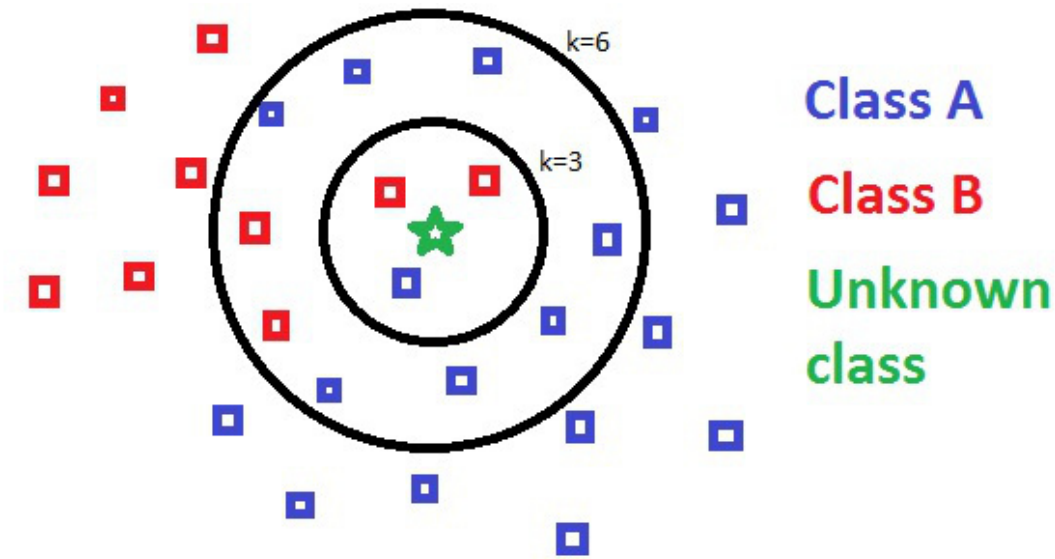

Figure 1: KNN classifier algorithm[8]

- For k=3 class A
- For k=6 class B

Implementation

User-User Based Collaborative Filtering[4]:

- It selects a neighbourhood of users who are similar to that particular user and best represent his taste.
- Similarity is computed through various correlation and then scoring various items by those neighbourhood of users to generate a recommendation for him.
- Different similarities in user-user based collaborative filtering:

1. Cosine Similarity
2. Euclidean Distance
3. Pearson Coorelation

Implementation

Cosine Similarity:

- The cosine similarity[7] between two vectors (or two users on the Vector Space) is a measure that calculates the cosine of the angle between them.
- Cosine Similarity will generate a metric that says how related are two users by looking at their behaviour in rating the products.

$$sim(A, B) = cos(\vec{A}, \vec{B}) = \frac{\vec{A}.\vec{B}}{\| \vec{A} \| * \| \vec{B} \|} \tag{1}$$

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \tag{2}$$

Euclidean Distance:

- Euclidean distance[3] is also known as simply distance. When data is dense or continuous, this is the best proximity measure.
- The Euclidean distance between two points is the length of the path connecting them.The Pythagorean theorem gives this distance between two points.

$$distance(i,j) = \sqrt{\sum_{i \in item} (s_i - s_j)^2} \tag{3}$$

$$DistanceBasedSimilarity = \frac{1}{1 + distance(i,j)} \tag{4}$$

Implementation

Pearson Coorelation:

- Correlation between sets of data is a measure of how well they are related. It shows the linear relationship between two sets of data.
- Pearson's correlation coefficient [1] when applied to a population is commonly represented by the Greek letter ρ (rho) and referred as the population correlation coefficient .

The formula for ρ is:

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} \tag{5}$$

where:

- cov is covariance
- σ_x is standard deviation of X
- σ_y is standard deviation of Y

Implementation

Popularity Model:

- Popularity model recommends the most popular products rated by the users.
- It is the most trivial and simple form of recommender system and doesn't consider any behaviour of user.
- We have recommended top 10 most popular products to every user.
- This model is very less flexible and intrusive to users.

Result Analysis

- Building a testing framework before implementing all these algorithms.
- Each data in the test dataset is a triplet of product id(asin), customer id, and rating.
- The product and customer are sent to our recommendation system for predicting, and the rating is our ground truth for testing.
- The score is measured with Accuracy, Mean Squared Error (MSE), precision and recall. After picking the test dataset, our test system returns an average MSE for the dataset.

Result Analysis

Calculation of similarities among items based on summary review of the product using k-nearest neighbour algorithm.

Neighbors	Accuracy	M.S.E.	Precision	Recall
1	53.84	68.67	.59	.54
5	61.89	50.19	.59	.62
9	64.77	44.64	.59	.65
13	67.23	40.74	.61	.67
17	68.38	39.19	.62	.68
21	68.81	38.43	.62	.69
25	68.63	37.63	.59	.69
29	69.43	36.73	.61	.69

Table 1: Performance of Item based recommender for different value of neighbours

Result Analysis

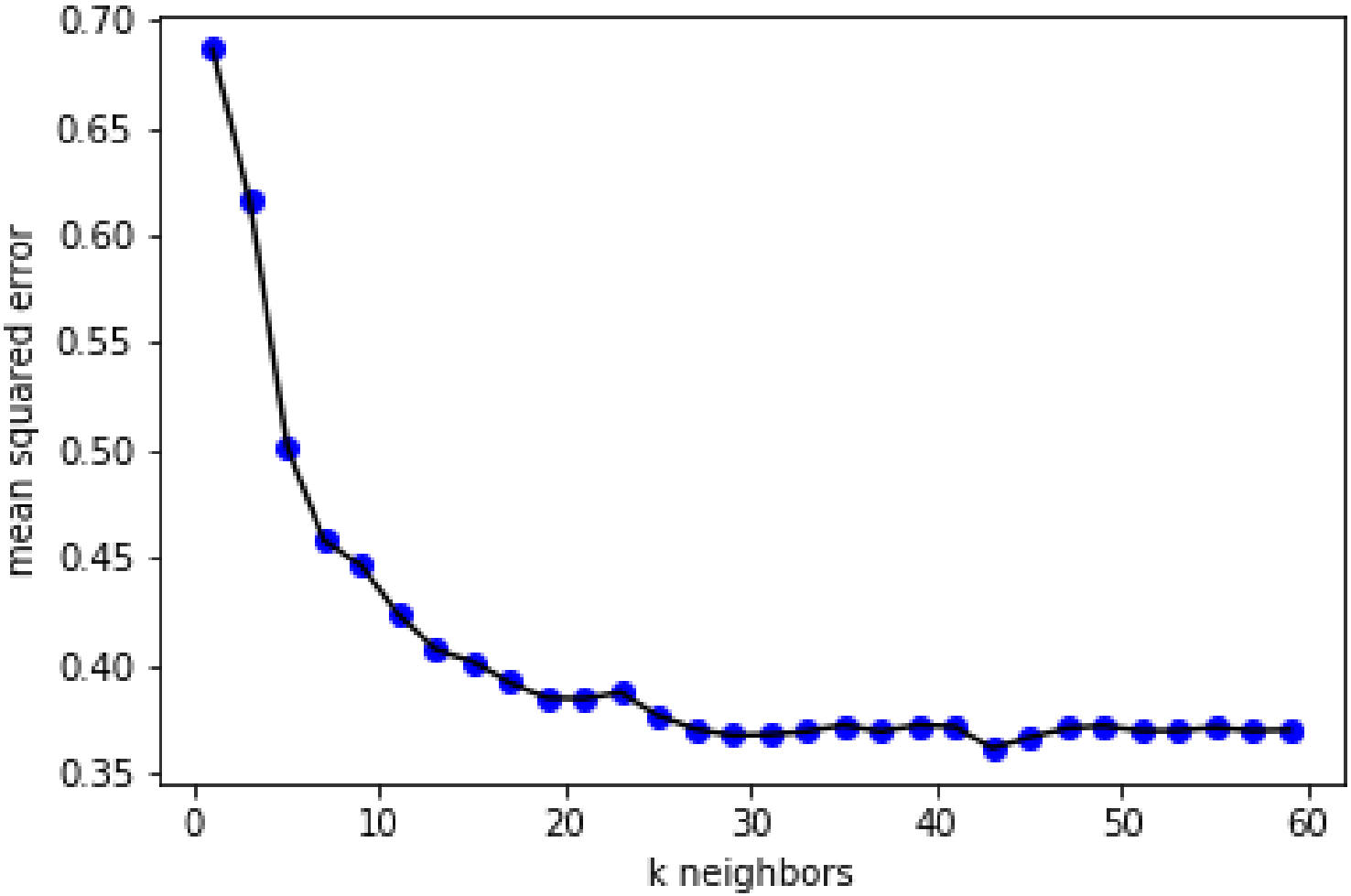

Figure 2: MSE vs K neighbours

Conclusion

- Recommendation systems help users discover items they might not have found by themselves and promote sales to potential customers, which provide an effective form of targeted marketing by creating a personalized shopping experience for each customer.
- In terms of effectiveness measured with mean squared error (MSE), for all users, Item-Item Similarity has the best result, then followed by User-User Similarity, and Popularity Model is the worst.
- Lots of companies have such kind of systems, especially for e-commerce companies like Amazon.com, an effective product recommendation system is very essential to their businesses.

Future Work

- We plan to improve the effectiveness and performance by exploring a hybrid system which will apply different algorithms on different user segments.
- We also need to experiment on different criteria to decide whether a user is a new user or an old user, and then choose the criterion that has the best result.

References


Real statistics using excel: Correlation: Basic concepts.
[Online; accessed 1-April-2018].


Badrul Sarwar, George Karypis, J. K.-J. R.
Item-based collaborative filtering recommendation algorithms.
ACM (2001), 285–295.


Elena Deza, M. D.
Encyclopedia of Distances.
Springer, 2009.


Guanwen Yao, L. C.
User-based and item-based collaborative filtering recommendation algorithms design.
ACM (2001).


Julian McAuley.
Amazon product data.
[Online; accessed 15-March-2018].


Ruining He, J. M.
Modeling the visual evolution of fashion trends with one-class collaborative filtering.
Ups and Downs 4, 2 (2016), 507–512.


Singhal, A.
Modern information retrieval: A brief overview.
Bulletin of the IEEE Computer Society Technical Committee on Data Engineering 24, 4, 35–43.


Yun-lei Cai, Duo Ji, D.-f. C.
A knn research paper classification method based on shared nearest neighbor.
3–4.

Thank You!