INCORPORATING SERENDIPITY INTO CONTENT BASED RECOMMENDER SYSTEMS USING DBSCAN ALGORITHM

Srishti Nigam[1], Rohit Kumar[2]

[1,2]*Department of Computer Science, University of Petroleum and Energy Studies, Dehradun, India*

ABSTRACT

With the number of choices and options continuously increasing on the internet, need for filtering systems such as Recommender Systems is also at its peak. Recommender systems help filter the content according to the user's or item's past history of preferences. Content based filtering is one such technique which recommends items based on a comparison between the content of the items and a user profile. However it faces the problem of lack of serendipity which prevents the user from experiencing new, fresh and fortunate results varying from their past preferences. This paper analyses this problem and aims to acquaint content based recommendation systems with serendipity with the help of DBSCAN clustering algorithm. We have used movie recommender system for implementing and explaining our solution.

Keywords: *clustering, content-based, DBSCAN, recommender, serendipity.*

I. INTRODUCTION

The two main types of recommender techniques are Content based filtering and collaborative filtering. There is also a third type known as hybrid filtering which is a combination of the above two. Content based filtering is based on the idea that a user is likely to have similar level of interest for similar items. In content-based filtering technique, recommendation is made based on the user profiles using features extracted from the content of the items the user has evaluated in the past[1][2].

Collaborative filtering on the other hand, works by building a database (user-item matrix) of preferences for items by users and matching users with relevant interest and preferences by calculating similarities between their profiles to make recommendations[3].The idea behind hybrid techniques is that a combination of algorithms will provide more accurate and effective recommendations than a single algorithm as the disadvantages of one algorithm can be overcome by another algorithm [4]

However, Content based filtering faces several issues due to lack of serendipity. We can understand serendipity as coming across a fortunate result while looking for something completely unrelated to it. A user biased towards action movies might be missing out on a highly-rated comedy movie as it was not recommended to him by the recommender system. A traveller usually going for holidays on beeches and mountains might prefer a desert safari tour, if recommended to him.

In this paper we aim to look into the problem of lack of serendipity and find out a solution to this problem using DBSCAN (density based spatial clustering of applications with noise) clustering algorithm. We have used DBSCAN to take advantage of the concept of finding noise value in DBSCAN. DBSCAN is very helpful for anomaly/noise detection. Also it can find arbitrary shaped clusters as opposed to only circular clusters. Here is how DBSCAN works[5]:

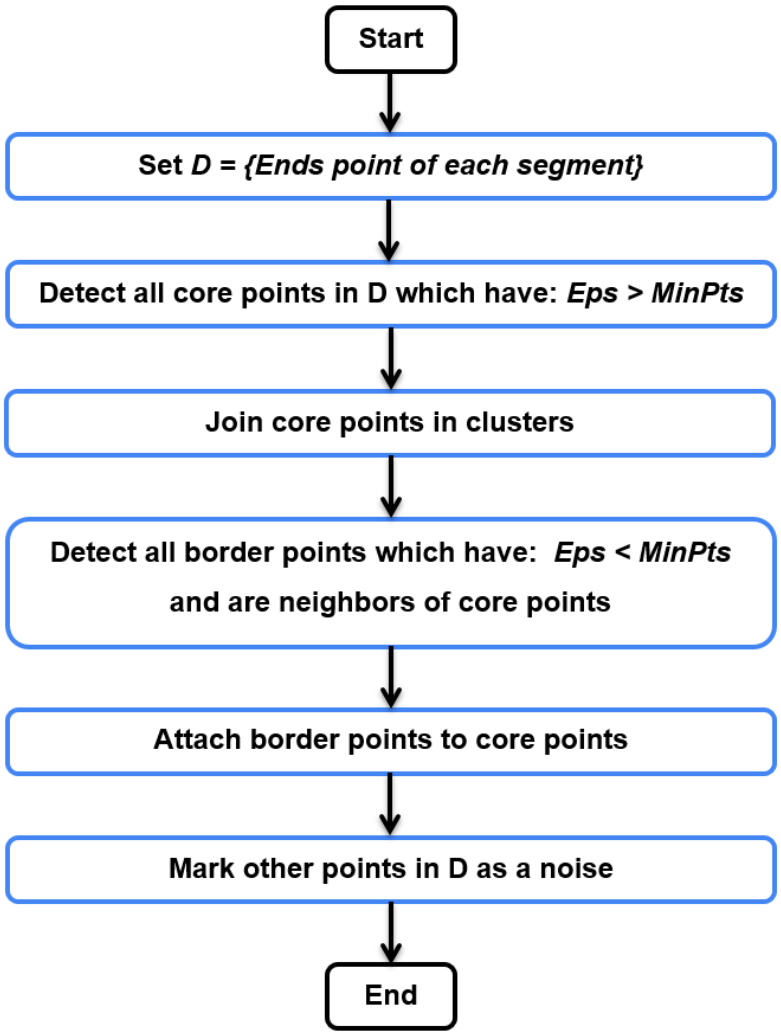

Fig 1: working of DBSCAN clustering algorithm

II. METHODOLOGY

- The methodology used in acquainting content based recommender systems with serendipity involves DBSCAN clustering algorithm. We take advantage of the concept of finding noise value in DBSCAN.
- Here we explain the process by taking the example of a movie recommender system with four possible categories of movies: Action, Comedy, Horror and Thriller. These category clusters have been already defined with the help of DBSCAN.
- Let us assume that a user usually watches Action movies and is an Action fan.

Fig 2: Serendipity in Content Based Recommender System

- Now a movie C belonging to the comedy category and having a rating r_{movie} is considered for Recommendation.
- We have some pre-set values of epsilon (max distance) and min-points (minimum number of points for a point to be a core point in a cluster).
- We therefore apply DBSCAN to the cluster of movies' ratings already belonging to the comedy category also including the ratings of movie C in the cluster.
- After applying DBSCAN, we get the updated Comedy cluster. Now if the movie C is a not noise point in the updated comedy cluster, the movie is not recommended as it lies in the cluster and is ordinary, Thus it is not really worth watching. But if it is a noise point, we proceed with the next step.
- We find the mean rating of the updated Comedy cluster 'mean_$r_{cluster}$' and compare the r_{movie} with mean_$r_{cluster.}$
- If $r_{movie} > mean\_r_{cluster}$, it implies that the comedy movie C is actually better rated than the other average comedy movies present in the Comedy cluster. Hence movie C is exceptional and is worth a watch. It will be suggested to the user.
- Else, we do not recommend the movie C to our user

II. CONCLUSION

Sometimes by chance, we stumble upon things which we really like even though they are not even remotely close to what we were searching for. Since content-based recommender systems make recommendations based on the user's past preferences, they were unable to make these fortunate, chance recommendations. Thus we have now incorporated serendipity into content-based recommender systems with the help of DBSCAN and overcome their previous disadvantage. This introduces users to a whole new set of possible recommendations, of which they were previously deprived. This solution can now be applied to e-tourism recommender system, book recommender system, technical paper recommender system, etc.

REFERENCES

[1] Burke R, Hybrid recommender systems: survey and experiments. User Model User-adapted Interact 2002;12(4):331–70.

[2] Bobadilla J, Ortega F, Hernando A, Gutie´rrez A, Recommender systems survey. Knowl-Based Syst 2013;46:109–32.

[3] Herlocker JL, Konstan JA, Terveen LG, Riedl JT, Evaluating collaborative filtering recommender systems. ACM Trans Inform Syst 2004;22(1):5–53.

[4]Schafer JB, Frankowski D, Herlocker J, Sen S, Collaborative filtering recommender systems. In: Brusilovsky P, Kobsa A, Nejdl W, editors. The Adaptive Web, LNCS 4321. Berlin Heidelberg (Germany): Springer; 2007. p. 291–324. http://dx.doi.org/10.1007/ 978-3-540-72079-9_9.

[5] El Bahi, Hassan & Zatni, Abdelkarim, Document text detection in video frames acquired by a smartphone based on line segment detector and DBSCAN clustering. Journal of Engineering Science and Technology. 2018;13.