International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 6 Issue 1, November-December 2021 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Study of Recommendation System Used In Tourism and Travel

Shikhar

Lovely Professional University, Phagwara, Punjab, India

ABSTRACT

This study is based on Recommendation Systems and its Types used in Tourism and Travel Website. Recommendation Systems are used in websites so that it can recommend item to a user based on his/her interest and on the basis of user profile. In this paper, I design a recommender system for recommending tourist places based on content based and collaborative filtering techniques. This method combines both behavioural and content aspects of recommendations. The flow for the research is that first of all using cosine similarity, weighted ratings and Location APIs we build a content-based system. The process is carried out by comparing the features of the item with respect to the user's preferences. Then followed by collaborative filtering techniques such as Correlation and K-Nearest neighbour in which items predict the interest of the user on an activity considering the evaluation that a particular user has given to similar activities.

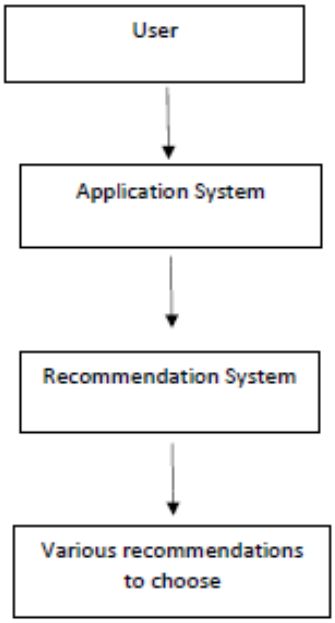
KEYWORDS: *content based, collaborative filtering, cosine similarity, kNN, Correlation*

How to cite this paper: Shikhar "Study of Recommendation System Used In Tourism and Travel" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-6 | Issue-1, December 2021, pp.1028-1033, URL: www.ijtsrd.com/papers/ijtsrd47922.pdf

IJTSRD47922

I. INTRODUCTION

The Recommendation Systems are a technological feature for social process and it's a way of suggesting like or similar items and ideas to a user's specific way of choosing and thinking. The very basic idea of recommendation systems is to automate the explicit filtering done by users, so that users are comfortable and at ease while choosing their options. Motivation for recommender systems come from some quotes like "I like this book; you might be interested in it", "I saw this movie, you'll like it", "Don't go to see that movie!", "I visited this wonderful place you also visit it", "This place has got negative vibes I don't recommend going!". In today's world business are switching towards recommender systems for improving their sales. Users may find new books, new places to visit, new movies, new music, new food, new dress which was previously unknown to them. It also provides people with opposite recommendation which will not be liked by the user. Large e-commerce websites use this tool to suggest items a consumer may want to purchase. The most common ways to use recommendation systems are using past purchase querying, genre matching, search style algorithms and surveys filled out by past users for the use of new users. The basic concept of recommendation system is given below: -



II. Recommender System Types

There are basically three types of recommender systems: -

A. Collaborative based Recommender Systems

It is also known as social filtering system which uses aggregation of user's preferences and recommendations to other users based on similarity in behavioural patterns.

B. Content based Recommender Systems

In this type there is use of supervised machine learning used to induce a classifier to discriminate between interesting and uninteresting items for the user.

C. Knowledge based Recommender Systems

In this system knowledge about users and products used to reason what meets the user's requirements, using discrimination tree, decision support tools and case-based reasoning (CBR)

In this paper we basically use the first two techniques i.e., Collaborative and Content based techniques for recommending tourist places and places of interest to user which is going to be discussed in detail in later part of the paper.

III. Problems Associated In Building Recommender Systems

A. Lack of Data

The most time consuming and very genuine problem related to recommendation systems are lack of data as effective recommendations could not be possible otherwise. The e commerce giants having excellent recommendation systems solely rely on the efficient data collection and massive storage they have for their data. For example, Amazon, Netflix, Google, Spotify etc all have very large database and based on that they have their fantastic recommendation system working.

B. Ever Changing Data

Nothing is static in the information and data world. Every second new data is generated and useful information gained as a result algorithmic approach will find it difficult to keep up with this trend. There are also more attribute adding on a certain item which results in failing of algorithms and produce undesired recommendations.

C. Sparsity Problem

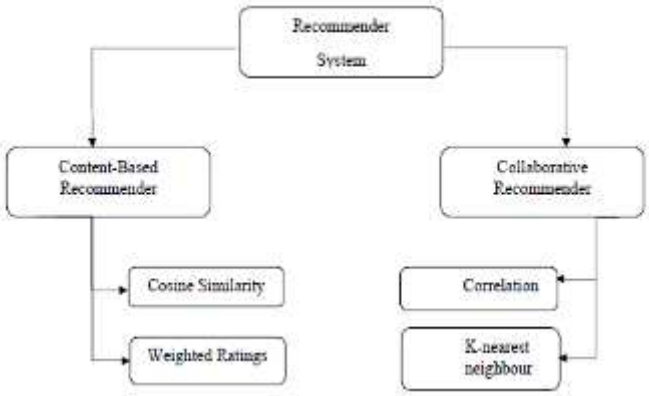
Sparsity issue emerges because of sparsity of rating grid for example client thing rating framework. This issue emerges because of utilization of the framework without giving rating and criticism for the framework. Despite the fact that we have numerous clients utilizing the suggestion framework yet we have not very many appraisals from those clients about various things in which they enjoyed or even detested. A few clients give bogus appraisals as they might suspect rating isn't really significant and they arbitrarily give evaluations thus these rating go into the suggestion calculations and afterward we get off-base results.

D. Cold Start Problem

Cold start issue emerges when no data is found with regards to a thing or the client in the suggestion framework. The virus start issue emerges in two situations – first in the event that we don't have any data about the new client who is getting to the framework. The motor can't suggest the initial time clients as there is no client profile assemble, second situation is the point at which another thing is presented which is extremely extraordinary and unknow to other people. For instance, assuming you visited a far off area which you believe merits thinking about a vacationer area of interest yet is obscure to other people. For this situation the suggestion motor doesn't prescribe that spot to explorers as they pass up on this chance. To stay away from cold beginning issues, we can utilize content-based arrangements and different AI methods.

IV. Method

In this paper I have used Content Based and Collaborative Recommender System to recommend places of interest to the users. Here you can see the flow diagram for building the recommendation system: -



A. Content Based Filtering Technique

In content-based filtering the technique that is used to recommend is "the items to users which are almost alike to the ones that the user wished or desired in the past". It is done by building a relation between item and its properties in term of Matrix then select the most similar items to the target item by computing similarity based on the features associated with the compared items using various mathematical functions. The common similarity function that used is Cosine similarity.

The two important techniques used in content-based filtering technique are: -

1. Cosine Similarity: -

The mathematical definition of cosine similarity states the angle between two n-dimensional vectors in an n-dimensional space. It is the dot product of the two vectors divided by the product of the two vectors divided by the product of norm of the two vectors (magnitude).

Cosine similarity is computed using the formula:

$$similarity(A,B) = \frac{A \cdot B}{\|A\| \times \|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \times \sqrt{\sum_{i=1}^{n} B_i^2}}$$

The values ranges from -1 and 1 , where -1 is perfectly dissimilar and 1 is perfectly similar.

Here in this recommendation system we can utilize this comparability to work out similitude between two spots. For instance, to get a vacationer place proposal dependent on inclinations of clients who have given comparative evaluations to different spots of interest they have seen.

The technique included registers closeness between all sets of things. The cosine comparability takes after balanced calculation implying that the come about because of registering the likeness of Item A to Item B is same as figuring similitude from Item B to Item A. Consequently, we can hence register the score for each pair of hubs once. We don't process similitude of things themselves.

Cosine likeness can be determined over just over non invalid aspects. The cycle hope to get same length of records for all things. Any other way, longer records will be managed to the length of the most brief rundown. This is a significant angle since information we get from the web are not spotless and contains a great deal of invalid qualities. We really want to clean the information prior to applying this capacity in order to acquire wanted outcome.

.Let us take an example to see how it works,

Suppose there are two vectors X and Y given as: -

X = (5,0,3,0,2,0,0,2,0,0) and Y = (3,0,2,0,1,1,0,1,0,1). To know how similar X and Y are we can use cosine similarity formula and calculate the similarity score.

Dot product of X and Y: X·Y = 5*3 + 0*0 + 3*2 + 0*0 + 2*1 + 0*1 + 0*0 + 0*1 = 25

$\|X\| = \sqrt{(25+0+9+0+4+0+0+4+0+0)}$

$\|Y\| = \sqrt{(9+0+4+0+1+1+0+1+0+1)}$

Similarity (X, Y) = 0.94

2. Weighted Ratings: -

The weighted ratings/scoring approach is to derive quantitative value for each competing item on your list, here is the cases of producing weighted scored of tourist destinations given in the dataset. We can use these values to determine which items the recommendation systems can prioritize so as to recommend the specific according to the weighted score, so as to push these scores in the efficient content-based algorithms so as to recommend items.

Mathematically we must define some parameters. To begin with we must first define how we will evaluate each potential item. For example, we may use the increase usage, increased revenue, increased trial.

For each of those, you need to set the "weight then". Depending on the stage of the business, that may change. We can take an analogy of weighted average to understand this concept of weighted ratings as follows: -

Suppose you're taking test and your teacher says that homework will make 40% of the score, tests will make 50% of the score and quizzes will make 10% of the score. In order to calculate the weighted average under those terms, you'll first calculate the average in each category. Let's say you get an average of 91% in homework, 89% in tests and 84% in quizzes.

Now first divide each by 100 to convert in decimal form which gives you homework: 0.91, tests: 0.89, quizzes: 0.84

Next multiply each category by its appropriate weighting factor, expressed as a decimal. Since homework is 40% of the score, you'll multiply homework category with 0.4 and similarly 0.5 with test category and 0.1 with quiz category so we get

Homework: 0.91*0.4 = 0.364

Tests: 0.89*0.5 = 0.445

Quizzes: 0.84*0.1 = 0.0084

Now add the results together: 0.364+0.445+0.084 = 0.893

This is the weighted score and in percentage form it is: 0.893*100 = 89.3%

Here I have explained you a simple example to understand the meaning of weighted ratings or scoring. Similarly, implementation has been conducted for various tourist places with different categories and their rating to calculate overall weighted rating so as to obtain the best recommendation according the content and data.

This is an effective strategy for scoring the places based on weighted scoring so as to we can enhance the result and effectiveness of cosine similarity and also pre-processing also included.

B. Collaborative Based Filtering Technique

The basic idea or we can say the main problem behind collaborative filtering technique is "trying to predict the opinion the user will have on the different items and be able to recommend the best items to each user"

The procedure for collaborative filtering involves trying to predict the opinion the user will have on the

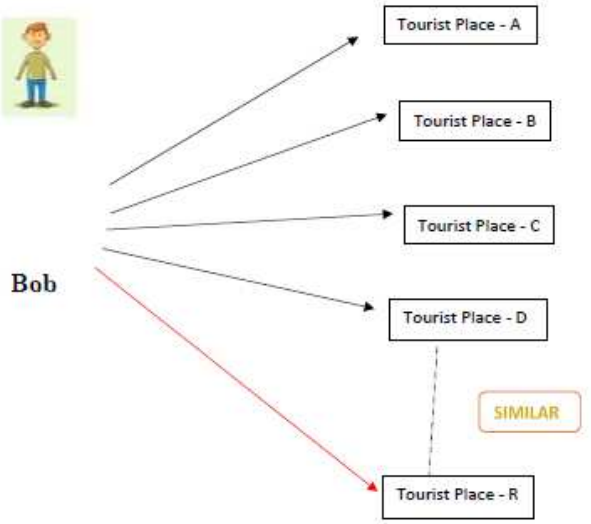
different items and be able to recommend the "best" items to each user based on the user's previous likings and the opinions of other like-minded users.

Here in tourism and travel website for recommending the place of interest to the users there should be already data present for m users and a list of n items.

Memory based collaborative filtering would be useful in developing a systematic and effective recommendation system. The general method to utilize the entire user-item database to generate a prediction. In order to do so we need to develop a model of user ratings.

The biggest challenge for collaborative algorithms is the Sparsity problem-evaluation of large item sets, user's purchases are under 1%. It is difficult to predict based on nearest neighbour algorithms. There is also an issue of scalability in real time tours prediction as nearest neighbour requires computation that grows with both the number of users and the number of items.

The main technique uses in this research paper for recommending tourist places is the Item based Collaborative Filtering algorithm which looks into the set of items the target user has rated and computes how similar they are to the target item and then selects K most similar items.



To understand Item Based Collaborative Filtering let us take an example. Suppose our user Bob wants to choose a place for travelling with his family. The job of recommendation system is to provide him with a new place based on his past preference. We will first search the places visited by Bob and let's call these places to be A, B, C and D where he has previously visited. Now we will search places similar to the above 4 places. Suppose we find a place R similar to place D, there is highly likely chance that Bob will also like the place R because it is similar to one Bob has already visited. Hence, we will suggest the place R to Bob.

So basically, item based collaborative filtering is all about finding items similar to the ones user has already liked.

Now many questions arise like How to find similar Items? And what if there are multiple similar items then in that case which item to suggest first? To answer these first we need to understand the intuition behind the process and a detailed mathematical analysis behind the Item Based Collaborative Filtering.

To find similarity between places/items we can use the cosine similarity technique to find the how similar the items on the basis of score obtained by items. Cosine similarity works but it doesn't take into account of the optimistic behaviour of users. Different users can rate the same places/items differently depending upon how optimistic they are. On the scale of 5, one could rate the place/item 5 while another could rate 3 even though they both very much liked the place/item. To account this, we make a small change to our similarity formula discussed earlier in cosine similarity section.

We subtract the user rating of a given item/place with that user's average rating which normalizes rating to the same scale and helps overcome optimism issues. We call it adjusted cosine similarity.

There is also another method where instead of subtracting user's average rating we subtract place's average rating. This tells us how much given user ratings deviate from average place/item rating. This technique is known as **Pearson similarity**. Here we will confine to cosine similarity for calculating the similarity between places.

1. K-Nearest Neighbor (KNN)

KNN is the go-to mode for Item Based Collaborative Filtering and provides a good platform for building recommender systems. This method uses database in which data points are separated into several clusters to make inference for new samples.

Since we are using Cosine Similarity for find similar places/items and KNN relies on this item feature similarity. When KNN inference about a tourist hotspot, it will calculate the "distance" between target place and every other place present in the database, then it ranks the places and returns the top K nearest neighbor of the most similar places to visit by a user.

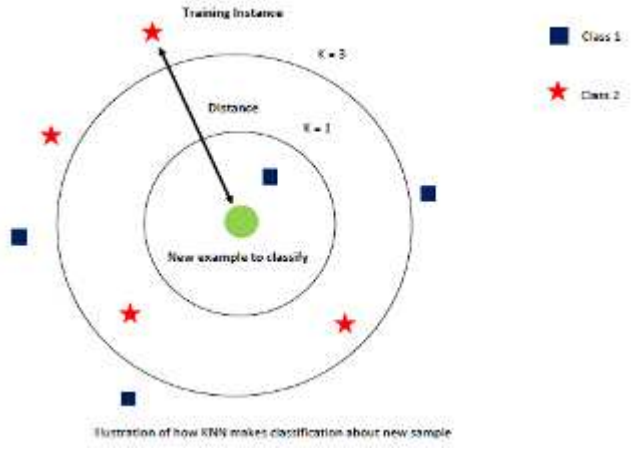


Illustration of how KNN makes classification about new sample

V. Similar Work

In present-day research the Recommender System utilized various areas to channel data, which has been broadly taken advantage of in business, elevating items and administrations to clients. Recommender System is a customized device that gives clients a rundown of accessible things that are appropriate to their person. The proposal says totally dependent on testing where we will pose inquiries to the client and note the appropriate response. After that, we will apply the suggestions to our framework based on criticism given by clients. To effortlessly assemble the right excursion by fulfilling travel inclinations, a gathering profile will be transcribed by a travel service. Also, settling on the right outing requires needs time on the grounds that the quantity of deterrents considered and the circumstance is significant. In the previous year, the Recommender System has changed significantly as far as time. When all is said and done, will be one of the significant supporters of their present setting. Proof from late examination proposes that is about the individual and area-based Recommender framework with uncommon reference to guest's space, talks about plan and execution issues to convey area based the travel industry related substance administrations.

VI. Research Gap

At whatever point a guest visits an obscure spot, we attempt to look at the best destinations accessible, there site and furthermore attempt to investigate nearby strength food sources and items accessible on those destinations.

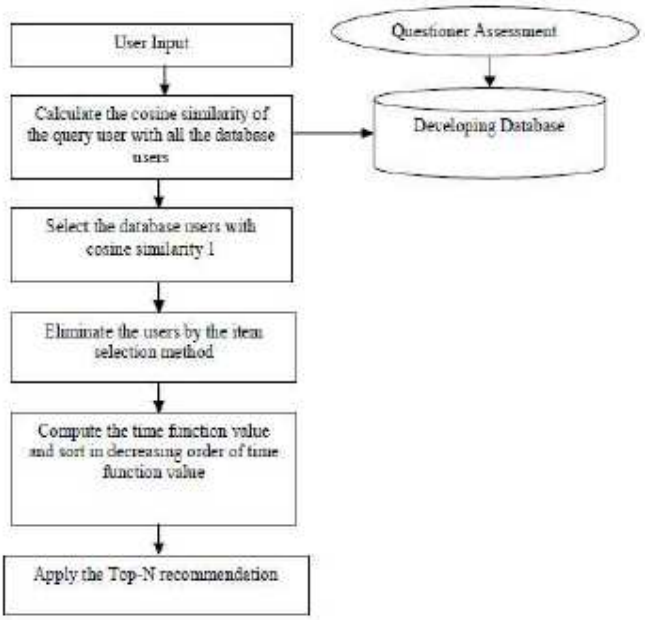
Some of the time a guest needs to take the assistance of different local people to track down destinations, and food too the items found on those destinations as the ones that are generally famous with them. Be that as it may, once in a while the language might be an obstruction to powerful correspondence with the neighborhood individuals. Along these lines, this paper proposes a method for suggesting it driving locales and food and items accessible on those destinations to guests when a guest is intrigued to go while booking lodgings on the web. The guest ought to give some data about the spot, the food and item inclinations while booking a ticket and the framework suggests high-n locales in a hurry famous where the guest can track down those sorts of food and items. In different books are created to prescribe locales to clients who utilize Collaborative separating technique. However, the proposed calculation utilizes a collective separating calculation that utilizes neighborhood clients with the end goal of suggestion. Nearby clients are those clients of their present area stay in the match with the's guest region and those clients have effectively visited the scene. The framework ought to likewise have the option to prescribe the furthest down the line highlights to the client so the client can utilize those proposals The proposed framework utilizes shared separating and time capacity to have the option to create the most recent suggestions for the client.

VII. Proposed Model

The proposed model starts with a test where the site is moved up to observe similar clients for the given info question client gave. It utilizes a collective sifting strategy for proposal when utilizing cosine-based similarity to observe indistinguishable clients for a specific visitor asking client. It utilizes the perspectives on nearby individuals about the different locales in the space that have been visited by those spaces. It estimates their perspectives about the site as far as the rating esteem additionally characterized by specific boundaries, for example, swarm the executives, security and site cleanliness. The sort of site the guest needs to visit might be profound or verifiable, or intriguing or a position of amusement. Nearby clients likewise give data about neighbourhood forte food sources and items accessible on the site at rating rates. The consumer weighs the food according to factors such as cleanliness, taste and price. Products are rated by local users on terms such as quality and price. It sees the same users based on cosine similarity value. The system computes the cosine similarity of the querying user with all the users present in the database by using the following formula.

$$similarity(A,B) = \frac{A \cdot B}{\|A\| \times \|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \times \sqrt{\sum_{i=1}^{n} B_i^2}}$$

The entire experimental design involved in the whole system can be represented below: -



All hypertext links and section bookmarks will be removed from papers during the processing of papers for publication. If you need to refer to an Internet email address or URL in your paper, you must type out the address or URL fully in Regular font.

VIII. Conclusions

The recommender system technique is used in a website for recommending places of interest for tourist in India and implement accordingly.

Acknowledgment

I would like to thanks my mentor Professor Dipen Saini and my group for helping me do this research on recommendation system.

References

[1] Chiu, D. K., & Leung, H. F. (2005). Towards ubiquitous tourist service coordination and integration: a multi-agent and semantic web approach. In Proceedings of the 7th international conference on Electronic commerce (pp. 574-581). ACM.

[2] García-Cumbreras, M. Á., Montejo-Ráez, A., & Díaz-Galiano, M. C. (2013). Pessimists and optimists: Improving collaborative filtering through sentiment analysis. Expert Systems with Applications, 40(17), 6758-6765

[3] Aggarwal, C. C., Han, J., Wang, J., & Yu, P. S. (2004). A framework for projected clustering of high dimensional data streams. In Proceedings of the Thirtieth international conference on Very large data bases-Volume 30 (pp. 852-863). VLDB Endowment

[4] De Maio, C., Fenza, G., Gallo, M., Loia, V., & Parente, M. (2017). Social media marketing through timeaware collaborative filtering. Concurrency and Computation: Practice and Experience.

[5] Ding, Y., & Li, X. (2005). Time weight collaborative filtering. In Proceedings of the 14th ACM international conference on Information and knowledge management (pp. 485-492). ACM.

[6] Fenza, G., Fischetti, E., Furno, D., & Loia, V. (2011). A hybrid context aware system for tourist guidance based on collaborative filtering. In Fuzzy Systems (FUZZ), 2011 IEEE International Conference on (pp. 131-138). IEEE.

[7] Garcia, I., Sebastia, L., Pajares, S., & Onaindia, E. (2011). The Generalist recommender System GRSK and Its Extension to Groups. Web Information Systems and Technologies,215-229.

[8] Hui, T. K., Wan, D., & Ho, A. (2007). Tourists' satisfaction, recommendation and revisiting Singapore. Tourism management, 28(4), 965-975.

[9] Cosine Similarity, Neo4J Graph Data Science Library, https://neo4j.com/docs/graph-data-science/current/alpha-algorithms/cosine/

[10] Collaborative Filtering for Recommender Systems, https://ieeexplore.ieee.org/document/7176109