Name: Greeshma Pillai

Student\_id: R00182989

HYBRID RECOMMENDATION SYSTEM USING KNN

INTRODUCTION:

Nowadays, many of the online platforms show suggestions on what to watch, which song to listen, may know a person, you may also like etc. These are the recommendation system filtering a vast amount of data available online as per the interest of any particular user. Recommendation systems are an increasingly prominent part of the web, accounting for up to a third of all traffic on several of the world’s most popular sites.[1]

Earlier every other item were sold in stores and stores had limited shelf spaces so only a fraction of items were available to the customer but on the other hand an online store can accommodate millions of such items. The distinction between the physical and on-line worlds has been called the long tail phenomenon. [2]

Into Thin Air and Touching the void:

An example of how long tail and recommender system influenced events. In 1988, Joe Simpson wrote a book named Touching the Void. Though reviews were good, it was only a modest success, and soon was largely forgotten. Then, a decade later Jon Krakauer's Into Thin Air, another book about a mountain-climbing tragedy, became a publishing sensation. Suddenly, Touching the Void started to sell again.

Booksellers began promoting it next to their Into Thin Air displays, and sales continued to rise. In early 2004, IFC Films released a docudrama of the story, HarperCollins released a revised paperback. By mid-2004, Touching the Void was outselling Into Thin Air more than two to one. In short, Amazon.com recommendations. The online bookseller's software noted patterns in buying behaviour and suggested that readers who liked *Into Thin Air* would also like *Touching the Void*. People took the suggestion, agreed wholeheartedly, wrote rhapsodic reviews. More sales, more algorithm-fuelled recommendations, and the positive feedback loop kicked in.

This created the *Touching the Void* phenomenon by combining infinite shelf space with real-time information about buying trends and public opinion. The result: rising demand for an obscure book.[3]

HYBRID RECOMMENDATION SYSTEM:

To overcome the disadvantages of the Content based filtering and Collaborative filtering (basic types of recommendation system), a hybrid method can be used. Hybrid systems combine two or more recommendation strategies to benefit from their advantages. Here a combination of recommendation system and machine learning algorithm such as Nearest Neighbors is used. Most businesses probably use hybrid approach in their production recommender systems.[4]

METHODOLOGY:

Aim: Comparing Recommendation System using KNN with Euclidean distance and Cosine similarity

Dataset:

The dataset is used from MovieLens. 25 million ratings and one million tag applications applied to 62,000 movies by 162,000 users. Includes tag genome data with 15 million relevance scores across 1,129 tags. Released 12/2019.[5]

Data are contained in: genome-score.csv, genome-tags.csv, links.csv, movies.csv, ratings.csv, tags.csv

K-Nearest Neighbors:

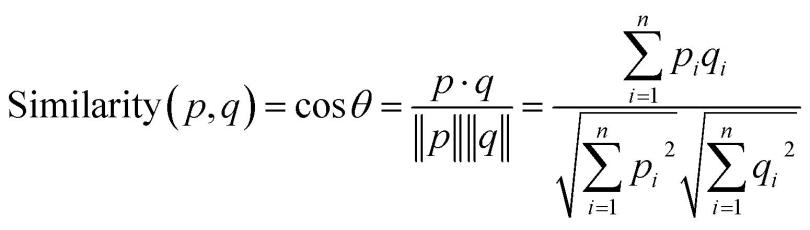
KNN is a non-parametric, lazy learning method. It uses a database in which the data points are separated into several clusters to make inference for new samples. KNN does not make any assumptions on the underlying data distribution but it relies on item feature similarity.[4]

Euclidean Distance:

[Euclidean](http://en.wikipedia.org/wiki/Cosine_similarity) Distance is a metric used to measure how similar the two items or documents are. It measures the distance between two vectors projected in multi-dimensional space. Here, less the distance more similar are the items. The distance between two vector is calculated using Pythagoras theorem.

Cosine Similarity:

[Cosine similarity](http://en.wikipedia.org/wiki/Cosine_similarity) is also used to measure the similarity but here size of the document doesn’t matter. It measures the cosine of an angle between two vectors projected in multi-dimensional space. Mathematically, the cosine of the angle of between two vectors is derived from the dot product of the two vectors divided by the product of the two vectors’ magnitude.



Cosine of -1 shows that two items are dissimilar and 1 shows that two items are completely similar.[6]

CONCLUSION:

The recommender systems will continue to play an important role be taking into consideration the amount of data available, which thereby will guide as many users to take better decisions.

Over years of development of personalization and recommendations, Netflix has reduced churn by several percentage points and saved more than $1B per year.[6]

Similarly , Amazon’s recommender system increases sales by [around 8%](https://dl.acm.org/citation.cfm?id=2764488) [1]. Several [other studies](https://dl.acm.org/citation.cfm?doid=2365952.2366014) of recommender systems also consistently report increases of several percentages in page views, time spent on page and revenue .

REFERENCE STYLE: IEEE

REFERENCES:

[1] A. Sharma, J. M. Hofman, and D. J. Watts, “Estimating the Causal Impact of Recommendation Systems from Observational Data,” in *Proceedings of the Sixteenth ACM Conference on Economics and Computation - EC ’15*, Portland, Oregon, USA, 2015, pp. 453–470, doi: 10.1145/2764468.2764488.

[2] A. Rajaraman and J. D. Ullman, “Mining of Massive Datasets,” p. 340.

[3] C. Anderson, “Long Tail : Why the Future of Business Is Selling Less of More,” p. 282.

[4] “Prototyping a Recommender System Step by Step Part 1: KNN Item-Based Collaborative Filtering.” [Online]. Available: https://towardsdatascience.com/prototyping-a-recommender-system-step-by-step-part-1-knn-item-based-collaborative-filtering-637969614ea. [Accessed: 25-Feb-2020].

[5] “MovieLens,” *GroupLens*, 06-Sep-2013. [Online]. Available: https://grouplens.org/datasets/movielens/. [Accessed: 21-Feb-2020].

[6] “MachineX: Cosine Similarity for Item-Based Collaborative Filtering - DZone AI,” *dzone.com*. [Online]. Available: https://dzone.com/articles/machinex-cosine-similarity-for-item-based-collabor. [Accessed: 29-Feb-2020].

[7] F. Barros, “The Netflix Recommender System: Algorithms, Business Value, and Innovation.”