*International Journal of Engineering & Technology*, 7 (2.6) (2018) 126-129

International Journal of Engineering & Technology

*Website: www.sciencepubco.com/index.php/IJET*

***Research Paper***

# A Comparison of recommendation algorithms based on use of linked data and cloud

**Ankita Ranjan [1]*, Vinay M [2]**

*[1]Department of Computer Science, Christ University, Bengaluru, India*
*[2] Department of Computer Science, Christ University, Bengaluru, India*
**Corresponding author E-mail: ankita.ranjan@mca.christuniversity.in*

## Abstract

Recommendation generation is a critical need in today's time. With the advent of big data and the increasing number of users, generation of most suitable recommendation is essential. There are many issues already associated with recommendations such as data acquisition, scalability, etc.. Moreover, the users today look to get best recommendations at the minimum effort on their side. Thus it becomes difficult to manage such huge amount of information, extract the needed data and present it to the user with least user involvement. In this research, we surveyed some recommendation algorithms and analyze their applications on an open cloud server which uses linked data to generate automated recommendations.

***Keywords***: *Recommender system; Linked data; Cloud Service*

## 1. Introduction

In this era, data available to all users are abundant. It becomes difficult to retrieve the kind, amount, and type of data that is needed by the user. A recommender system thus acts as an information filtering system, and it helps to filter the information that is already retrieved or filter the information even before retrieval. In general, we see there are two basic types of recommender systems- service and expert-oriented. Service oriented recommendation systems provide a recommendation to users for services such as movies, music, news, books, etc. On the other hand, expert recommender provides recommendation related to specific fields like restaurants, banks, financial services, etc. Each system has a different type of algorithm applied to generate the most suitable recommendations. The algorithms are extensive in nature and categorized as Content-based Collaborative, Hybrid. Apart these there are others called Knowledgebase, Demographic, and Utility, etc.

## 2. Ease of Use

### 2.1. Linked Data

Coming to our existing recommender systems we find that all of these recommender systems have a unique database. This database is created and maintained individually by all recommender systems and data is extracted as per use. So no two recommender systems can share each other’s database. This restriction leads to unnecessary clogging and redundancy of information. The solution to this problem is Linked Data.

Linked Data refers to [7] the process by which the whole web is connected. It means that through s set of typed data, which acts as links the whole web, is connected. This can apply databases found in different organizations at different geographical locations as well data that belong to one organization. The main underlying aim of linked data is to connect or bring together data that were not connected or lacked inter-operability. It’s interesting to note that linked data is explicitly defined, data published on web that is machine readable. Linked data not only has the capability to be turned into links, it can also be linked to other external dataset. According to the author [7] there are set of rules that define the guidelines in which data is published on Web such that all of the data published can be linked together to form a single dataset that has global access. These guidelines are:

1. Instead of name of things using Uniform Resource Identifier or URIs.
2. These URIs should be linked to HTTP to ease searching for them.
3. Whenever a URI is searched we need to make sure that the look up is done through the set standards such as RDF, SPARQL etc.
4. Linking URIs in order to let a user discover more things.

Linked Data is the process of which data is published on the internet in such a structured format so that it is linked one way or the other to all other data present. In other words, linked data suggests mechanisms for publishing information in a manner that all other data available can be related to it. Use of linked data will help to create a common general database specific in the genre. It will contribute to control the data clogging and redundancy problem.

### 2.2. Cloud Computing

Now even if we do not connect the all the data available on web but we create a single unified dataset for all recommender system we might solve the problem of redundancy.

But that will lead to new issues as to how the data can be accessed, should there be a common portal that allows all the contributing recommender system to access each other's data, moreover if there is such a portal where can we launch it, should it be based in a single geographical location or multiple. More questions like this arise when we link the data. The solution comes here in the form of Cloud Computing.

Cloud Computing [9] is a model that enables provisioning of services that heavily rely on the use of internet. This provisioning of services is dynamically scalable and can also be of virtualized services. Most common Clod Services include office application like word processors, spread sheets, presentations etc. Nowadays nearly all big , high earning software corporations such as Amazon, Microsoft, Google, IBM etc. are providing with cloud services over a varied range of applications and devices.

Cloud servers can be modelled as an

- Infrastructure like services through which it's available for rent to organizations, and they access the storage space through private or public networks.
- Platform as a service where the whole system can be built upon the server and access is universal.
- Model cloud as Software where the application or the processing happens on a cloud whenever the user accesses the server through an API or Application Programming Interface. of 0.5cm between columns.

## 3. Recommendation Algorithms

### 3.1. Content Based

A content-based recommender system as the name suggests content or information about a user to recommend an item.

It focuses on trapping a user's past preferences and analysing them to find what is the personal taste of the user. This means that for every user using a particular site, the recommender system traps all data of user on that site. Then this data is analysed to find a pattern that is common to the user and gives us an insight to his/her habits and taste. Like the kind of movie a user should watch next is recommended upon the ratings he/she has given to any movie watched by them before. It checks the genre, actor, director or any other vital element and compares it with movie database.

This type of recommendation system faces the problem of ramp up. Ramp up problem states that the recommendation cannot be made for an item that user has no previous rating or preference. For example, if a user who searches music, movies, etc. suddenly asks for a book recommendation it becomes difficult for content-based systems to suggest a book as the user has no history of book reading or rating.

The only disadvantage here is the uncertainty of finding data about an item that user has never searched before or an item which is new in the market.

### 3.2. Collaborative Filtering

This type of Recommendation algorithm uses the user's peer's interests to recommend items.

This recommender system looks not at the user's taste/preference but groups people based on the products they like. If two users like/buy/view the same product then the two users are grouped together and called peers. Then either of the user gets recommendation of an item which was not viewed or liked by him/her but was liked/viewed/bought by that peer. That is a user is recommended an item that is rated high by another user who has shown similar interests to the user in the past. For example, a user A buys or rates an item X on an online shopping site, another user B buys or rates the same item X later on the same site.

Now when the user A goes to shopping site again, he/she will be suggested with items that were rated high or bought by user B. This is done because the system sees the similarity in taste thus suggests the item one user has liked before other.

This brings an end to the ramp up the problem as the user can search for elements that are entirely new to them and by the interests of peers sharing the same taste, or peers of peers to suggest if the item is entirely new.

Now, when this algorithm is applied to a recommender system A, a new User U using the system for the first time faces the cold start problem. The recommender system is unable to suggest anything as the user has no peers, no previous common taste, etc. It becomes a huge problem when the database is unique. However, again if the user is entirely new to the environment, it makes it tough to suggest anything for the user.

There are many types of Collaborative filtering methods such as:

- Item based CF: Based on ratings of some item by two different users.
- User based CF: Prediction of user's rating based on another user's rating of same item.
- Memory based CF: Predictions by performing operations on available data.

### 3.3. Hybrid Filtering

Hybrid Filtering as the name suggests is the process by which two or more different filtering algorithms are combined with each other to provide the best recommendations. The system uses both the user's personal data as well as the other techniques of taste and preference comparisons to give out best-suited suggestions.

For example, a system using Content and Collaborative together will use both the user's past interests and the preference of users like him/her to generate a list of recommendations that are best suited.

This leads to one problem of maintaining a huge database of one user. Each user and his all personal details; his previous selections, etc. are kept along with his/her general preferences peer. This all takes a toll on the system and creates problems relating time, accessing issues, etc.

### 3.4. Knowledgebase Recommender System

This type first tries to understand the user and his attributes along with the product and its characteristics. Then it uses this information to find out the customer product need relationship. It tries to establish a need relation for a product then only recommends it to the user. Now it eliminates the ramp up problem but has a huge disadvantage as it needs an engineered database and an ever updating set of personalized recommendations.

### 3.5. Demographic System

The demographic study such as gender, class, creed, geographic location, age, etc. is studied in order to create a user profile. This user profile not recommends for individuals but provides a generic recommendation based on demographic. Like an elderly user would like a different set of products than a teenager. Thus in this precise personal recommendation cannot be found but it gives a general overview of likes and dislikes.

Now considering all the systems of recommendation we find that hybrid system is most flexible and has least disadvantages among others. Now if all the recommender system in the world is hybrid systems and all have an individual database of each user. Considering the user's preference all these databases have similar information, thus if hosted on a server even cloud server, the redundant data will lead to clogging and hamper performance.

## 4. Related Work

In the past, linked data related work has been done to analyze and in detail study regarding Collaborative Recommender Systems in the field of linked data [2]. Here using the same dataset and comparing the results of the applied algorithm with and without the use of linked open data. Some studies have been also done to see how content based recommender systems work in the same linked data environment [6]. For this the authors took different movie datasets and then used them to generate Linked data based recommendation. Other related areas have analyzed the working of linked data and the history, present and past [7], where the different application, publishing tool etc. related to linked data were clearly analyzed, whereas [8] outlined details about how data published on web, giving a preface to linked data. On the other hand the past work related to cloud are more in depth having an analysis of Remote Venue Cloud based Recommendations [4].
Apart these many authors [7] [10] have analyzed the different recommendation algorithm based on their own parameters. While how Hybrid filtering is a better approach than the other algorithms [6] have been described in detail by comparing hybrid algorithm with all other algorithms along with analyzing which of the already known hybrid type of algorithm from weighted, switching, mixed etc. is the best combination for recommendation generation. But all of these recommender systems which use linked data have never had a cloud based server and any recommender system not using linked data is backed up by a cloud base. Thus we are suggesting a comparison that analyzes the advantages of combining these two.

## 5. Comparison of Algorithms

Content based filtering analysis based on Linked data by Mirzi, Noia, Ostuni[3], recommends movies based on different databases. The content based filtering sample values for data of hold-out testing 100 samples has been taken into consideration by Hun, Chong, Ong [4].
Considering the collaborative filtering based comparison as done by Heitman and Hayes [2], we can evaluate content based recommender with and without the use of linked data.
The hybrid filtering has used data of hold-out testing of 100 samples by Hun, Chong, Ong [4].The linked data based hybrid recommender systems that have been analyzed by Khrouf and Troncy[5].
Here Precision and Recall are used as measure to analyze the algorithms. Precision is the ratio of all the items that were liked by a user and recommended divided by number of products actually recommended, while Recall is the ration between the numbers of items liked and recommended with the total number of items that were liked by the user.

**Table 1:** Algorithms Comparison

| Algorithm Name | Comparative Measure | | |
|---|---|---|---|
| | Basis | Precision | Recall |
| Content- Based (without Linked Data) | Previous interest | 18.3% | 38.1% |
| Content-Based (with Linked Data) | Previous interest with use of Linked Open Data | 35.35% | 50.0% |
| Collaborative (without Linked Data) | Based on Peer Interest | 2% | 7% |
| Collaborative (with Linked Data) | Based on Peer Interest with use of Linked Open Data | 14% | 33% |
| Hybrid (without Linked Data) | Mix of two or more types of algorithms | 7% | 34.4% |
| Hybrid (with Linked Data) | Mix of two or more types of algorithms with use of Linked Open Data | 24.81% | 69.1% |

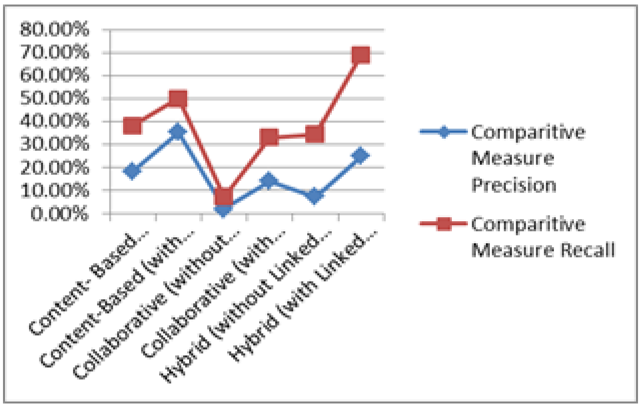

**Fig. 1:** Graphical Representation

The same results can be seen when these recommendation algorithms were put to test in the cloud background. The analysis done by Lee and Lee [6] shows the following results

**Table 2:** Cloud based Comparison

| Algorithm Name | Comparative Measure | | |
|---|---|---|---|
| | Basis | Precision | Recall |
| Cloud based Recommendation | All algorithms of recommender system | 45.81% | 46.55% |

## 6. Conclusion

The recommendation generation is a tiring process and leads to a collection of a large amount of mostly overlapping information. When these all information is stored at the main point of access, it creates many issues. Moreover, a database which is unique leads to uninformed or less suitable recommendation generation.

The different algorithms have different drawbacks and can't be combined with the best generation as this will result in additional data checking, filtering and the elaborate and complex process will be more time-consuming.

So when we put one or more algorithms in the Linked Data environment and remove the unique database concept, then redundancy of data is reduced, and individual problems with algorithms are sorted.

Now when we move this database to the cloud, it does not create the problems it would have if the database were on a single system. This means the problem of accessing data from a cloud reduces the storage issues as well as decreases the issues that come with remote access of the same database. Thus we can suggest a system which shares data from all other systems, is located on cloud and uses any the hybrid variety of recommendation as it is the most versatile and adaptable type.

## Acknowledgement

I would like to thank my guide Prof Vinay M to be a pillar of support through the writing of this paper. I would thank my institution to provide me this opportunity. Lastly, big thanks to god almighty, my friends, and family who gave continuous support and had faith on me.

## References

[1] Heitmann, Benjamin, and Conor Hayes. "Using Linked Data to Build Open, Collaborative Recommender Systems." AAAI spring symposium: linked data meets artificial intelligence. 2010.

[2] Huang, Zan, et al. "A graph-based recommender system for digital library." Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries. ACM, 2002.

[3] Khrouf, Houda, and Raphaël Troncy. "Hybrid event recommendation using linked data and user diversity." Proceedings of the 7th ACM conference on Recommender systems. ACM, 2013.

[4] Di Noia, Tommaso, et al. "Linked open data to support content-based recommender systems." Proceedings of the 8th International Conference on Semantic Systems. ACM, 2012.

[5] Ziegler, Cai-Nicolas, et al. "Improving recommendation lists through topic diversification." Proceedings of the 14th international conference on World Wide Web. ACM, 2005.

[6] Amini, Bahram, Roliana Ibrahim, and Mohd Shahizan Othman. "Discovering the impact of knowledge in recommender systems: A comparative study." arXiv preprint arXiv:1109.0166 (2011).

[7] Christian Bizer, Tom Heath, Tim Berners-Lee “Linked Data - The Story So Far” Proceedings of the Heath, T., Hepp, M., and Bizer, C. (eds.). Special

[8] Issue on Linked Data, International Journal on Semantic Web and Information Systems (IJSWIS). http://linkeddata.org/docs/ijswis-special-issue.

[9] Burke, Robin. "Hybrid recommender systems: Survey and experiments." User modeling and user-adapted interaction 12.4 (2002): 331-370.

[10] Ms. K. C. Sonawane1, Mrs. S. S. Ponde” A Survey on Cloud-Based Remote Venue Recommendation System” Proceedings from International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 4, April 2016