

SEMINAR REPORT

ON

# **Machine Learning with Social Media:**

## **Netflix**

**BY**

Shivani S. Shinde

T.E.Computer, Division – B

Roll No. : 3474

2017-2018

SEMINAR GUIDE

Mr. Hitendra Khairnar

CUMMINS COLLEGE OF ENGINEERING FOR WOMEN

KARVENAGAR, PUNE

Affiliated to Savitribai Phule Pune University

## **CERTIFICATE**

This is to certify that **Ms. Shivani S. Shinde**

has satisfactorily completed the seminar on

**“Machine Learning with Social Media: Netflix”**

in the partial fulfillment of her term-work (Seminar and Technical Communication Laboratory)  
as a part of syllabus for T.E. Computer Engineering in the Academic Year 2017-2018 as  
prescribed by Savitribai Phule University of Pune.

**Mr. Hitendra Khairnar**  
**Seminar Guide**

**Dr. Mrs. S.M. Kelkar**  
**Head of Computer Engineering Department**

**Dr. Mrs. M. B. Khambete**  
**Director**

## **PLAGARISM CERTIFICATE**

This is to certify that **Shivani S. Shinde**

has satisfactorily completed the seminar report on

**“Machine Learning with Social Media: Netflix”**

and the data present in this report is found to be 100 % plagiarism free.

This seminar report is towards the partial fulfillment of her term-work (Seminar and Technical Communication Laboratory) as a part of syllabus for T.E. Computer Engineering in the Academic Year 2017-2018 as prescribed by Savitribai Phule University of Pune.

**Name of the student**  
**and sign**

**Mr. Hitendra Khairnar**  
**Seminar Guide**

## **Acknowledgement**

I wish to express my sincere gratitude to respected Principal, Dr. Mrs.Madhuri M. Khambete and Head of Computer Department, Dr. Mrs. Supriya Kelkar for providing me an opportunity to present my seminar on “Machine Learning with Social Media: Netflix”.

I am highly indebted to my seminar guide Mr. Hitendra Khairnar for the guidance and constant supervisions. I would also like to thank her for giving me her precious time and encouraged me in completion of this report.

**Shivani S. Shinde**

**Roll No. 3474**

## Abstract

Artificial intelligence is nothing but a technology aiming to develop self-computing and self-understandable machines which are trained based on variety of data sets. Under the branch of Artificial Intelligence lies **Machine learning** and **Deep learning**. Machine Learning is a branch of Artificial Intelligence which aims at developing systems which do not have to be programmed explicitly for each and every use case of problems. These systems designed using machine learning, process the given data and learn on their own. Machine Learning is mainly categorized into two parts: Supervised and Unsupervised Learning. Google's self-driving car is the best example of Machine learning's supervised approach.

Machine Learning carries out its task using various algorithms. **Recommendation System** is one of the most used system in numerous fields right now. These use machine learning algorithms which provides end users relevant items or outcomes in various fields of their choice. Example of recommendation system can be online shopping sites like Amazon, Flipkart, Snapdeal, etc. These sits use recommendation algorithms to give their users items of clothing of their choice using the data of their previous purchase. These are self-learning algorithms which develop pattern or user choices and recommend similar items. Recommendation systems are applied in various fields like music, videos, shopping, medicines, agriculture, entertainment.

# **TABLE OF CONTENT**

**List of figures**

**List of Tables**

<b>Chapter 1</b>	<b>Introduction</b>	<b>1</b>
<b>Chapter 2</b>	<b>Description of Netflix</b>	<b>4</b>
<b>2.1.</b>	Netflix services	4
<b>2.2.</b>	How Netflix works	6
<b>Chapter 3</b>	<b>Various Algorithms in Recommender Systems</b>	<b>9</b>
<b>3.1.</b>	Types of recommendation approaches	9
<b>3.2.</b>	Decision Tree	12
<b>3.3.</b>	Training Algorithm	13
<b>Chapter 4</b>	<b>Algorithms in Netflix</b>	<b>14</b>
<b>4.1</b>	Restricted Boltzmann Machines Algorithm Matrix Factorization	14

<b>4.2</b>	<b>15</b>
------------	-----------

<b>Chapter 5</b>	<b>Conclusion</b>	<b>17</b>
------------------	-------------------	-----------

<b>References</b>	<b>18</b>
-------------------	-----------

## **List of Figures**

Figure No.	Title
1.1	Eliza System
2.1	Netflix's Working Service
2.2	How Netflix works
3.1	Collaborative Approach
3.2	Content-based Approach
3.3	Decision tree of movies
4.1	Artificial Neural network

## **List of Tables**

Figure no.	Title
3.1	Various Recommendation Algorithms



## Chapter 1: Introduction

**Social Media** is the new buzz word around the corner of technology. Since last decade from now, use of social media has increased tremendously. Digital medium like smartphones, tablets with a hand of internet has given birth to platform like Twitter, Facebook, Instagram, Gmail, YouTube has driven the communication of users like anything. Where Facebook and Instagram allow you to share posts in the form of pictures, videos, stories which last for 24 hours, having social handle like Twitter, allows you to express views in character limit of 240 (then, 140 characters).

Social media plays an important role in marketing of products. In early days like in 1970s or 1980s, publicizing a product or a company or an organization took huge efforts to pay for advertisement and one to one publicity. Organizations would appoint publicity heads for the same. Businesses now view social media as a powerful tool for digital marketing. Use of social media has completely revolutionized the way people look at marketing.

Not just marketing, social media helps people to communicate with other individuals no matter at what distance he/she is from oneself. Instant messaging service has increased the pace at which one can communicate with other. Facebook, Instagram, WhatsApp allow sharing of URLs which is key tool for promotion of various startups. Social media has been identified as best suited for all purposes of sending and receiving information through various mediums.

Communicating in early decades, before the birth of mobiles and internet, used to take place using hand written letters. These letters were transferred over to destination by a postman which used to take many days in transportation. Similarly, people needed to wait for days or even months to get replies in the form of letter. This method was completely overridden by, first, the telephones, then, smart phones and now, social media.

But it's not just the applications which been build on internet that has involved, the only source of buzz. It is the technique they use which helps social media become powerful as it is now. It

would not be wrong to make a statement that social media can implement telepathy in real life, for it knows what user seeks. This secret lies within the backend technologies used in various social media platforms. Few years ago, when social media started giving responses of user's choice, it felt like a magician knowing our desires and responding accordingly. However, people became aware of the actual technique used as and when the technology evolved. That was none other than **Machine Learning**.

Although machine learning is a common word now and may seem a 21<sup>st</sup> century terminology, it traces its roots back to 1950s to 1960s where first attempt of human machine interaction was made. ELIZA system which was programmed by Joseph Weizenbaum at MIT, was the first computer program which was appreciated in world of Machine learning. It is what we now call a "Chat Bot". The system used to interact with users and reply pretending to be a psychotherapist. Following was the first conversation of system ELIZA.

```
=====
      EEEEEEEE L      IIIIII ZZZZZZZZ      AAA
      E      L      I      Z      A      A
      E      L      I      Z      A      A
      EEEEE L      I      Z      A      A
      E      L      I      Z      AAAAAAA
      E      L      I      Z      A      A
      EEEEEEE LLLLLLLL IIIIII ZZZZZZ      A      A
=====
ELIZA > (PLEASE TYPE IN ALL CAPS) WHAT'S YOUR NAME DEAR ?
AMIT
ELIZA > HI, I'M ELIZA. WHAT DO YOU WANT TO TALK ABOUT ?
AMIT > CAN I TALK ABOUT MY PROBLEM ?
ELIZA > SURE... !
AMIT > DUE TO ON GOING EXAMS IN COLLEGE, I HAVE A LOT OF STRESS.
ELIZA > PLEASE RELAX A LITTLE. SLEEP WELL.
AMIT > THANKS FOR YOUR ADVICE.
ELIZA > NO MENTION.
AMIT > BYE
ELIZA > BYE AND KEEP IN TOUCH...
=====
```

*Figure 1.1 Eliza System [www.scaruffi.com]*

Machine learning is a powerful tool if used wisely. It is an approach where a user can program certain task for a computer and it must work on its own to carry out its task. For example,

Google's self-driving car is an excellent real time machine to demonstrate machine learning. The developers of this car have programmed it in such a way that user just needs to enter the location of place he wants to visit and the car drives towards the destination automatically. The car has been trained for several thousand of miles to learn when to increase or decrease the speed of car according to various terrains.

**Combination of social media and machine learning is the best suited application, NETFLIX.**

**Netflix** is an online entertainment portal where users can log in and watch their favorite TV shows, movies etc. This paid platform allows user access to free trial for a month. Netflix originated nearly around 1997 where it first provided delivery of DVDs of your purchase right into your home. While a decade after, in 2007, Netflix made its debut in online world of streaming. Stranger Things, Breaking Bad, The Walking Dead are available series to name a few. It allows user to have subscription plans for every three, six or nine months. Netflix allows downloading of selected tv shows and movies which can watched without having internet connection.

## **Chapter 2: Description of Netflix**

Reed Hastings and Marc Randolph founded a company which offered renting movies online, which were delivered at your home directly. Initially the company allowed users to watch unlimited movies in one-month subscription. Introduction and availability for computers increased in later 90s, and gradually the orders increased tremendously. Watching movie on the go, led to higher demand for ordering DVDs. As the technology advanced in 2000, the makers thought of getting digital. Thus, in 2007, an online movie streaming came into existence and this completely revolutionized the way people saw films. Watching movies of our choice at any time of the day, anywhere, increased number of viewers visiting the website day by day. Presently, Netflix has more than 70 billion active subscribers all over the globe and these figures increase on an average of 4-6 million subscribers daily each passing day.

### **2.1 Netflix Services**

So, what is it that made Netflix a huge success? Of course, availability of smart phones, computers, tablets have made streaming easier. But there's something more than this. It won't be wrong to state that Netflix does take care of its customer's requirements and fulfill them technically. In fact, Netflix has now mastered the art of knowing user's choice and recommending the type of movies and/or TV shows that users may like. Following diagram shows how Netflix achieved its services.



*Figure 2.1 Netflix's Working Service[[www.slideshare.net/adrianco/netflix-global-cloud](http://www.slideshare.net/adrianco/netflix-global-cloud)]*

Netflix first started with SaaS i.e. System as a Service, since it initially started home delivery of DVDs. This service was well appreciated by users which led the company move to build their own PaaS i.e. Platform as a Service.

Inculcating PaaS, Netflix provided an online platform for its users to watch movies or TV shows at their own pace and space. This rapidly made news since it was feasible for users. The company later joined hand with various cloud services like AWS Cloud, etc. to work productively. Following diagram depicts working of Netflix with integration of cloud services, customer's device and the platform.

## 2.2 How Netflix Works

### How Netflix Streaming Works

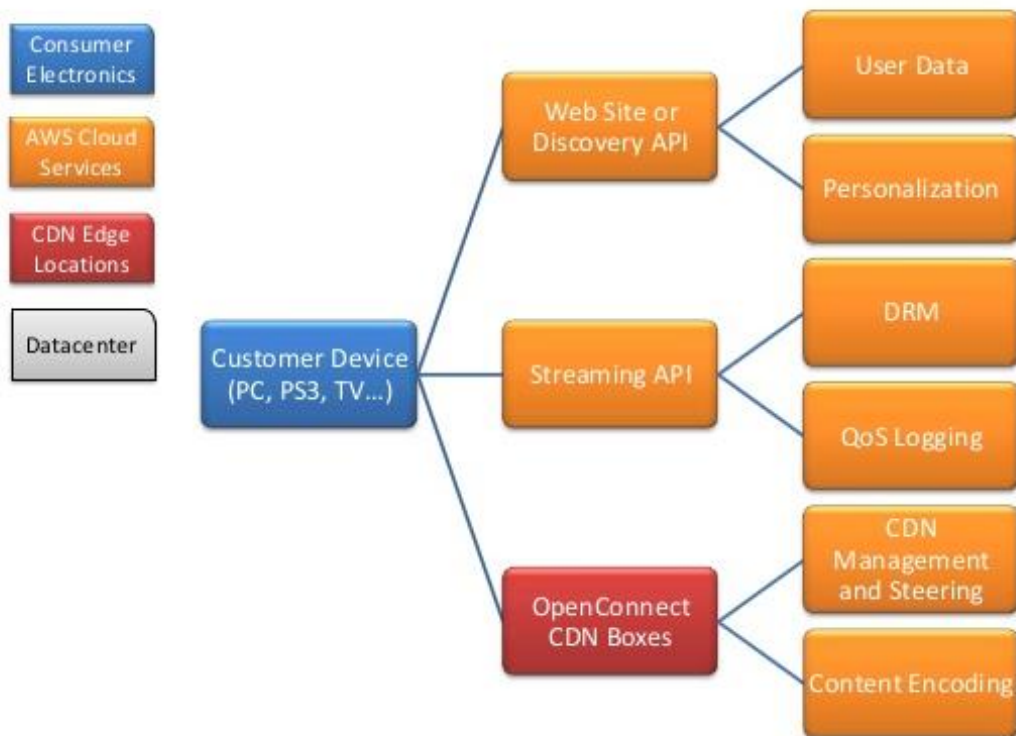


Figure 2.2. How Netflix works [<http://www.telco.com> Avi Dorfman]

The image shows how Netflix can be available on customer's device. It can be in the form of website on online platform or application on mobile operating platforms like Android or iOS. The user then connects to Netflix via using application or website. And then the user searches for the shows or movies he/she wants to watch. The searches which are made by user are stored in databases and content is returned in result set. User data is carefully stored and manipulated. This data is stored and used for future recommendations to the user.

## What Netflix does?

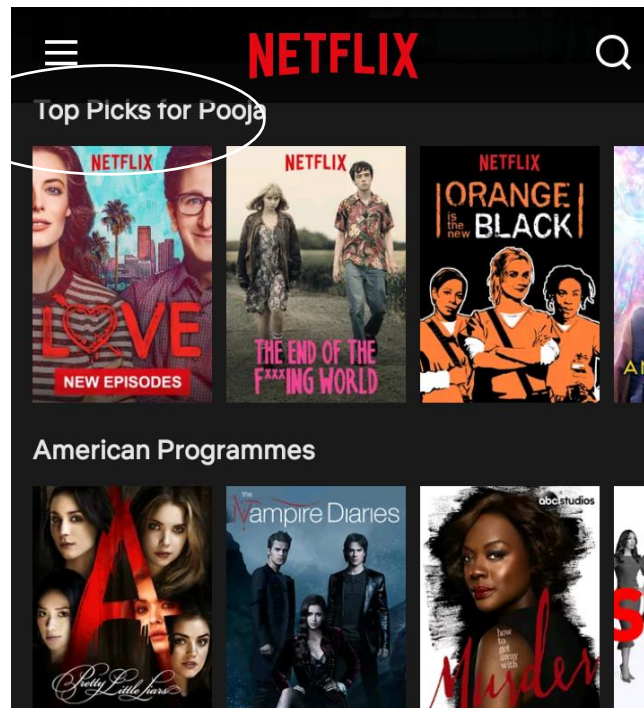
The user needs to login on to the website of Netflix, [www.netflix.com](http://www.netflix.com). The online portal displays various TV shows, movies. The user needs to subscribe monthly which sums around INR 500.

The user can then watch unlimited movies or tv shows within a month of subscription. The searches made by user are manipulated and the desired items are shown on screen.

### For example:

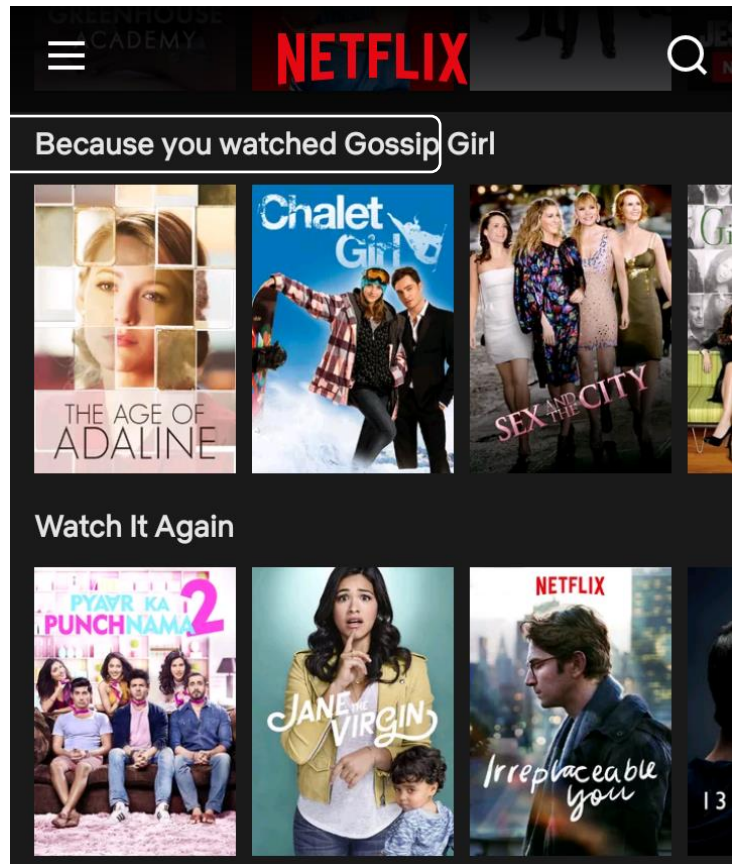
If a user watches one film of horror and two of comedy genre, the Netflix then recommends the user, the next movies based on their previous watch history. Suppose a user views a film related to teenagers, then Netflix recommends few movies based on the films which are made for teenagers. Following are the snapshots taken from the android-based application of Netflix.

*Here Netflix recommends a user named Pooja.*



Since a user watched a TV series

'Gossip Girl', similar shows are  
recommended.



Netflix 80% accurately recommends correct choices for user's next watch. But how does Netflix achieve that? The answer is **Recommender Systems**.

Various recommender systems are used when it's time to recommend a movie, songs, videos etc. These systems are nothing but algorithms which play crucial role in recommendation. Each algorithm has different use cases and scopes according to which a developer decides which algorithm will be best suited for his/her system/application.

Netflix also uses Recommendation algorithms. The use and selection of these algorithms has however changed in past ten years. These algorithms work on the input data which is previously watched movies of user and generate most likely to be watched movies as output. User reviews have shown that they have readily watched most movies or TV shows recommended to them on Netflix.



## Chapter 3: Various algorithms used in Recommender Systems

In order to generate correct output, various applications use an algorithm which is best suited for various use cases. Recommender systems aren't exception. Recommender system contains various algorithms which are used for different purposes. They use artificial intelligence to recommend items to users. Recommender systems were first used in early 1992.

### 3.1 Types of Recommendation approaches

Recommendation systems are broadly classified into following types:

1. Collaborative
2. Content based
3. Hybrid

#### Collaborative Approach:

In this approach the algorithm uses user data to process in for recommending an item or movie. For example, a user buys a music album. The information such as age, city, country, type of album purchased is all taken into account in this approach. Then it is compared with number of other users with the same information and the kind of music they bought is learnt and recommended to previous user.

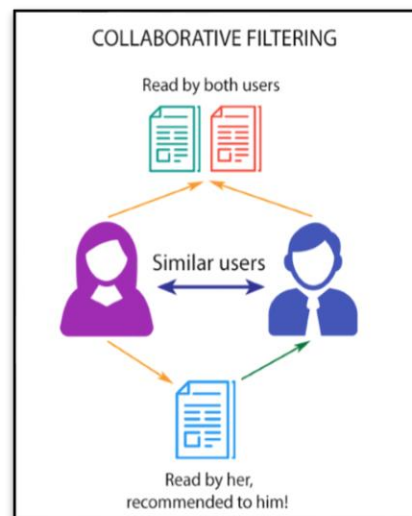


Figure 3.1 Collaborative Approach [(Source: The Marketing Technologist)]

Here, the recommendations are done by matching and comparing user data with one or more users.

### **Content based Approach:**

This approach based on characteristics of the items viewed by a user. For example, consider a user where he/she is looking for a smart phone. When a smart phone URL is accessed by the user on the browser, the algorithm collects all information regarding specifications of that smart phone like RAM, brand of smart phone, camera megapixels, processors, etc. Then the system browses more items from its database matching with those specifications. And then the other smart phones having same specifications re recommended to user.

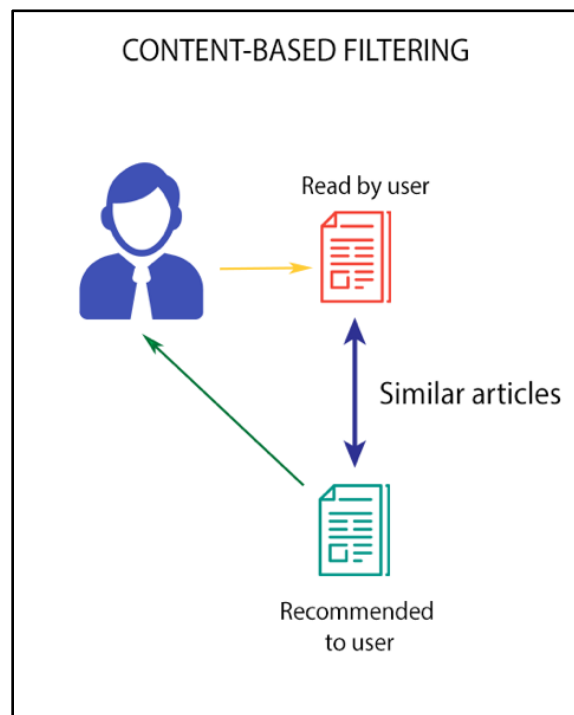


Figure 3.2 Content-based Approach [The marketing technologist]

## Hybrid Approach

As the name suggests, this approach is a hybrid of previous two approaches. That is these recommend items based on both; the user choices and the item data. For example, on a social network like Facebook, recommendation system may recommend profiles that are similar to the active user using collaborative filtering, by comparing their interests. And then, the system may consider the recommended profiles as items and then access their data to search for new similar profiles, which we call as Friend Suggestions, using content-based filtering.

Using collaborative or hybrid approach, the system must gather information about the user or the data items to recommend. This can be done implicitly or explicitly. Explicitly, user data gathering happens when users are fully aware that they are providing their information. For instance, when registering for any online website, users usually fill in a form that asks for their name, age, etc. Where as implicitly is what happens when the user browses online shopping store. The user does not know that the information of items they are browsing is being collected at the backend.

Following machine learning are mainly used in recommender systems, since machine learning forms an integral part of recommendation.

*Table 3.1 Various recommender Algorithms [The Use of Machine Learning: Algorithms in Recommender Systems: A Systematic Review]*

Category	Publications	Total
Bayesian	[20] [21] [24] [36] [39] [55] [66]	7
Decision Tree	[21] [36] [39] [55] [57]	5
Matrix factorization-based	[7] [53] [60] [61]	4
Neighbor-based	[25] [31] [39] [60]	4
Neural Network	[5] [21] [39] [41]	4
Rule Learning	[24] [25] [36] [58]	4
Ensemble	[10] [21] [57]	3
Gradient descent-based	[8] [14] [60]	3
Kernel methods	[14] [21] [66]	3
Clustering	[19] [20]	2
Associative classification	[36]	1
Bandit	[12]	1
Lazy learning	[21]	1
Regularization methods	[14]	1
Topic Independent Scoring Algorithm	[41]	1

Following are few algorithms which are used in recommender systems:

### 3.2 Decision Tree

Decision trees have been proved to be model based approach for recommendations. Using this approach, one can achieve efficiency and flexibility while handling various datatypes. These trees contain data or items to be recommended at leaf nodes. As and when the user browses data and likes or dislike an item, the tree internally learns the choices of users. The tree then separates out the characteristics of liked or disliked movies or items and modifies tree accordingly.

Consider an example of movie recommendation:

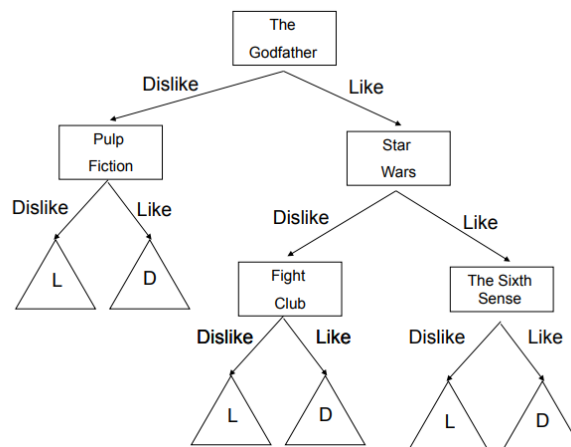


Figure 3.3 Decision tree of movies [Decision Tree Based Recommender System, by Amir Gershman, Amnon Meisels]

### 3.3 Training Algorithm

RBM's are trained to get the largest product of probabilities assigned to input values.

Equivalently it evaluates the probability for entire input of data sets.

The process for single sample can be done as follows:

1. Take a training sample  $v$  and a hidden activation vector  $h$  from this probability distribution.
2. Compute the *positive gradient*, outer product of  $v$  and  $h$ .
3. From  $h$ , collect  $v'$  of the visible units, then collect the hidden  $h'$  from this.
4. Compute the *negative gradient*, outer product of  $v'$  and  $h'$ .
5. Let the update to the weight matrix be the positive gradient minus the negative gradient, times some learning rate: .
6. Update the biases  $a$  and  $b$  analogously:

$$\Delta(a) = \epsilon(v - v'), \Delta b = \epsilon(h - h')$$

## Chapter 4: Algorithms in Netflix

Out of best suited algorithms using machine learning, Netflix predominantly uses two approaches which are :

1. Restricted Boltzman Machines
2. Matric Factorization

## 4.1 Restricted Boltzmann Machines Algorithm

In simple words, Restricted Boltzmann Machines algorithm is a fancy neural network that can learn on its own when provided with set of inputs. This is often used with collaborative filtering in Netflix. For the data set generation, around 8+ billion user's choices are considered.

For Netflix, the ratings of half million users who gave ratings from 1 to 5 on around 18,000 movies is given. These are the models which are generative which automatically detect the inherit pattern of gathered data by reconstructing their inputs. Instead of focusing on the right answer, it focuses on the data which nearly equivalent or closest to the data in given system.

Consider an artificial neural network which consist of visible or the input layer and a hidden layer.

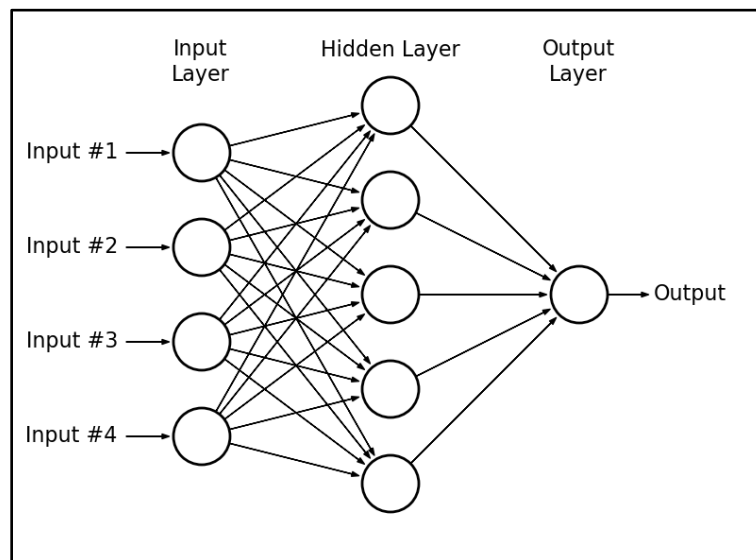


Figure 4.1 Artificial Neural Network [cstwiki.wtb.tue.nl]

The values of hidden layer must be found first in artificial neural network. Therefore, the data is forwarded from the input layer to hidden layer, which is then propagated backwards to the input layer. Comparison of the input values and original values is carried out. Weight and bias values are constantly read to get the values closest to the reconstruction. The only restrictions which applies is that no two nodes of input values should share any connection.

## 4.2 Matric Factorization

According to Netflix competition, matrix factorization is best proved technique over nearest-neighbor techniques with respect to product recommendations. This technique broadly focuses on enhancing user experience by recognizing patterns of user's choice. This model characterizes both items and users by vectors inferred from patterns of item rating. This method has become a favorite choice of developers for its accuracy and good scalability.

One of the strength of this model is if explicit data is not available i.e. if feedbacks, user ratings are not available, it utilizes implicit data. Many researchers have realized problems of huge data sets and sparsity which is the null values are stored in matrices since everyone does not rate the movies.

Suppose there are 943 users selected to rated say, 1682 movies then a matrix of  $(943 \times 1682)$  is formed. If we consider that on an average only 10% user rate movies, this leads to 90% sparseness.

If we have  $U$  set of users,  $I$  items, then let  $R$  be the matrix of  $|U| \times |I|$  which contains all the ratings of user. To find latent features, we need to find two matrices  $P(|U| \times K)$  and  $Q(|I| \times K)$  such that,  $R \approx P \times Q^T = \hat{R}$ .

## **Chapter 5: Conclusion**

It is thus stated that Netflix uses various recommendation algorithms in order to recommend best suited movies and TV shows. The algorithms are effectively used and applied which gives Netflix a success ratio of nearly 85%. The case study of Netflix is well studied. The recommender systems on scanning user data, produce most likely to be watched list of tv shows and movies.



## **References**

### **Papers:**

- 1) The Use of Machine Learning Algorithms in Recommender Systems: A Systematic Review by Ivens Portugal, Paulo Alencar, Conald Cowan.
- 2) Using Content-Based Filtering for Recommendation by NetlinQ Robin van Meteren and Maarten van Someren
- 3) A Decision Tree Based Recommender System by Amir Gershman, Amnon Meisels.
- 4) Item Based Collaborative Filtering Recommendation Algorithms by Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl.

### **Websites:**

- 1) [www.wikipedia.com](http://www.wikipedia.com)
- 2) [www.techleer.com](http://www.techleer.com)
- 3) [www.youtube.com](http://www.youtube.com)