Personalize News Recommendation System by Using Stack Auto Encoder

Ruchika Dive, P.R.Pote COE , Rohan Raut, P.R.Pote COE, Prof. V. B. Bhagat,
Sgbau Amravati, Amravati, University Amravati
ruchadive48@gmail.com, rohans.raut95@gmail.com, matevaishali2@gmail.com

Abstract-The popularity of Internet and mobile Internet, people are facing serious information overloading problems now a days. Recommendation engine is very useful to help people to reach the Internet news they want through the network. Recommender Systems have become the vital role in recent years and are utilized widely in various areas of social importance. In day to day life, users will not be able to read news every day due to heavy schedule. So to increase General knowledge of users we propose online recommendation systems which recommend news.

In existing system i.e. news delivery portals deliver popular news on home page of the portal but user's data according recommendation is not implemented yet. To overcome existing system problems we propose new recommendation system which automatically finds the news based on user's profiles. Using the stack auto encoder algorithm.

Keyword: *recommendation engine; collaborative filtering; stacked auto-encoder.*

*****

I. INTRODUCTION

The development of Internet and mobile Internet, more and more people use mobile devices to access the network. According to the data from CNNIC [2], in China alone there are almost 700 million people using mobile devices to access the internet. Every day many people use mobile to read news in their fast paced daily life. How to help the reader to reach the content as soon as possible is a very important issue not only to readers but also to news media [2].

There are many challenges for collaborative filtering tasks (Section 2). CF algorithms are required to have the ability to deal with highly sparse data, to scale with the increasing numbers of users and items, to make satisfactory recommendations in a short time period, and to deal with other problems like synonymy (the tendency of the same or similar items to have different names), shilling attacks, data noise, and privacy protection problems [3].

The development team tries to use the latest techniques to improve the algorithms, and improve the platform's service ability towards the users step by step. Deep learning is very helpful in its ability to learn the features and to do comparison. Stacked denoising auto-encoder, which has powerful characteristics of unsupervised learning and feature extraction ability, has been experimented to improve the platform's recommendation efficiency [2].

A. Recommender overview:

Recommender is an active research field in data mining and machine learning for some time. There exist different approaches for recommendation engines, such as

1. Content-based filtering,
2. Collaborative filtering, and
3. Hybrid techniques

Collaborative filtering, Collaborative filtering (CF) is one of the most important method. There are different type of CF method, such as neighborhood-based algorithms, model based algorithms and hybrid models [3, 4]. CF methods are described like this: the CF method is based on the phenomenon which includes collecting and analyzing a large amount of information on basis of users' behaviors, activities or preferences and predicting what users would like based on their similarity to other users. In CF, the information on personal preferences, tastes, and quality are all carried in (explicit or implicit) user ratings which can be utilized [5].

Neighborhood-based CF techniques, which can also termed as memory-based approach, in this technique the predication are made on the basis of the entire or a fraction of user's operation record, usually the user-item database in the system[4, 6]. Item-based CF currently is the most popular approaches. Neighborhood-based CF techniques can also be divided into user-based CF filtering, and item-based CF filtering [9]. This is used to make a prediction by evaluating the weighted average of all the ratings of the user or item on a certain item, or using a simple weighted average for the active user [3].

Model-based CF are formulate to make predictions using data mining, machine learning algorithms to find patterns based on training data. There are many model-based CF methods, including Bayesian networks, clustering models, latent semantic models, and newly RBM, Auto-Encoder based solutions. The memory-based CF and the model-based CF algorithms are coupled together as a hybrid method to make recommendation more effectively but the implementation of such approaches have increased complexity and expensive [7, 8]

B. Stacked Auto-Encoder:

Model-based approaches are those algorithms, which do the computing with predefined offline model, but presenting the result online. The design and development of models (such as machine learning, deep learning algorithms) allow the system to learn to recognize complex patterns based on the training data, and then make intelligent predictions for the collaborative filtering tasks for test data or real-world data, based on the learned models. There are several main stream model-based algorithms in use, such as Bayesian Net or Dependency Net algorithms , Clustering algorithms , Markov decision processes, Regression based algorithms, SVD based algorithms, and Auto-Encoder based algorithms. Model based algorithms have been investigated to solve the shortcomings of memory-based CF algorithms. Since deep learning has been a hot spot in machine learning, and improved the performance in pattern recognition and NLP greatly, it has been used in recommendation domain certainly. Stacked auto-encoder based CF is one of the deep learning approaches.

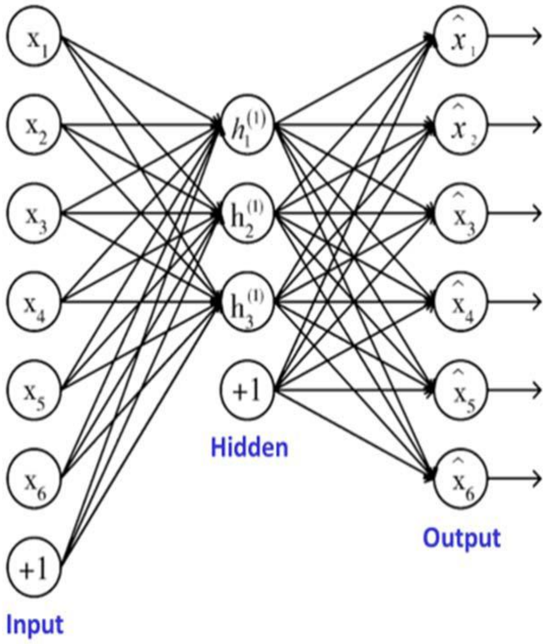

Fig. 1. The general structure of an Auto-Encoder

Now people are trying to use deep learning to improve the efficiency and performance of recommendation systems. Because of the powerful feature extraction ability of Stacked denoising Auto-Encoder (SDAE), it is the well-suited method in recommendation engine. Auto-encoder, which was mentioned by Rumelhart in 1986, is a kind of neural network and unsupervised learning methods which attempts to be trained to copy its input to output. It is typically a three layers neural network. The first layer is the input layer, the second layer is the hidden layer, and the third layer is the output layer, referring to Fig. 1. The mechanisms between the input layer and hidden layer could learn a dataset's low dimensional distributed representations from input layer's original high dimensional counterparts, just like an encoder. The mechanisms between the hidden layer and output layer just do the opposite things, which can reconstitute the original high dimensional information, just like a decoder. When using auto-encoder alone, the features extracted by the encoder are not robust enough. After adding the Gaussian noise, it is better. So for making the hidden layer to discover more robust features and to prevent it from simply learning the identity, denoising auto-encoder were created. It is a stochastic kind of the auto-encoder which does two things, trying to encode the input (preserve the information about the input), and trying to undo the effect of a corruption process stochastically applied to the input of the auto-encoder. Followed structure is a denoising auto-encoder :

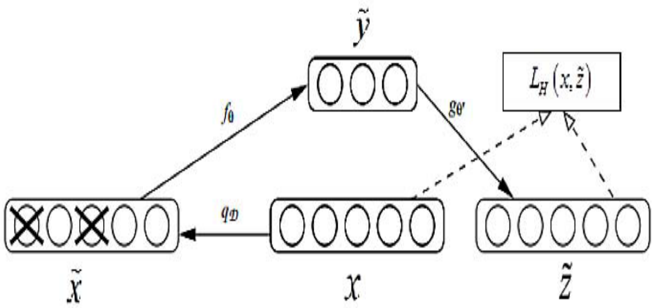

Fig. 2. The principle of denoising Auto-Encoder

A stacked auto-encoder is a neural network consisting of multiple layers of sparse auto-encoders in which the outputs of each layer is wired to the inputs of the successive layer. Denoising auto-encoders can be stacked to form a deep network by feeding the latent representation (output code) found on the layer below as input to the current layer[1].

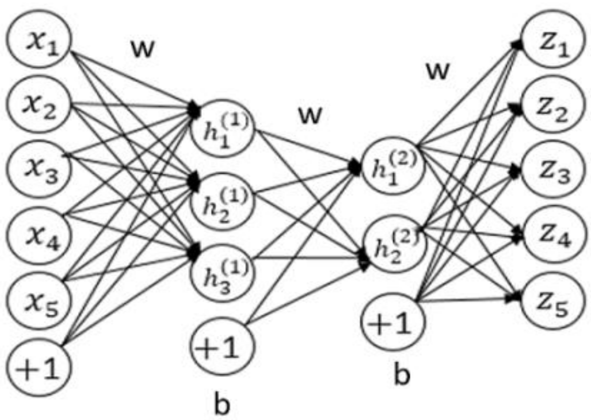

Fig. 3. The structure of a Stacked denoising Auto-Encoder

II. Conclusion

Recommender systems are designed to identify the item that a user will like or find useful based on the user's preferences and activities. Over the years, collaborative filtering, which derive these recommendations by leveraging past activities of groups of user, has emerged as the most prominent approach for solving this problem. Among the multitude of methods that have been developed, user and item based nearest-neighbor approaches are the simplest to understand and easily extended to capture different user behavioral model and types of available information.

III. References

[1] Sanxing cao, Nan yang "Online News Recommender Based on Stacked Auto-Encoder",May,24-26, 2017

[2] The 38th CNNIC Report, http:// www.cnnic.net.cn /hlwfzyj/hlwxzbg/hlwtjbg /201608/t20160803_54392.htm.

[3] X. Su, T. Khoshgoftaar, A survey of collaborative filtering techniques, Advances in Artificial Intelligence archive, 2009.

[4] J. Schafer1, D. Frankowski, J. Herlocker, S. Sen, Collaborative Filtering Recommender Systems, International Conference on Intelligent Systems Design& Applications, 2015, pp. 438-443.

[5] K. Yu, A. Schwaighofer, V. Tresp, X. Xu, H. Kriegel, Probabilistic Memory-based Collaborative Filtering, IEEE Transactions on Knowledge & Data Engineering, 2004, pp. 16(1):56-69.

[6] D. Goldberg, Using collaborative fi1tering to weave an information tapestry, Communications of the Acm, 1992,pp. 35(12):61-70.

[7] M. Pazzani, A framework for collaborative, content-based and demographic filtering, Artificial Intelligence Review, 1999, pp. 13:393–408.

[8] R. Burke, Hybrid Web Recommender Systems, Adaptive Web: Methods & Strategies of Web Personalization, Lecture notes on computer science, 2007, pp.4321:377-408.

[9] J. Breese, D. Heckerman, and C. Kadie, Empirical analysis of predictive algorithms for collaborative filtering, in Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, 1998.