

Influences of Deep Learning on Recommendation Systems

Neha R. Kasture

Department of Computer Application, Shri Ramdeobaba College of Engineering and Management, Nagpur, India
 Email: kasturenr@rknec.edu

Received: 20th September 2018, Accepted: 11th October 2018, Published: 31st October 2018

Abstract

Deep structured learning is a noteworthy advancement in the subset of machine learning. Recommendation not only requires domain knowledge but also needs data science intuition to cope up with the abundance and complexity of data in the age of information overload. The plethora of information available on internet and especially on social networking sites can be exploited to provide some personalized recommendations to the users. These recommendations could be predicting the next point of interest (POI) or stop over in the domain of tourism, judging the future preferences of the users from their past choices or predicting their socio-historical inclination from the data available through location-based social networks (LBSN's). Lot of research has been conducted recently which gives effective recommendation. The goal of this article is to provide a comprehensive survey and comparative analysis of the state-of-art research techniques based extensively on recommendation systems used for applications related to sequence learning.

Keywords

Deep Learning, Recommendation Systems, Location Based Social Networks, POI Recommendation, Sequence Learning, Social Media Analysis.

Introduction

Recommendation systems are a tool to filter the data of user's choice from dearth of online information. Such systems can ease the user's decision making process related to the products or services which they are interested in but have not seen from a huge array of available options. Lately recommendations are highly used in the tourism industry to predict the tourist's next stop over or point of interest based on spatial and temporal data like check-ins available through location based social networks (LBSN's). Recommendation systems have thus become the core area of industry to boost online websites and sales services with increased customer satisfaction. Deep learning is recently gaining remarkable attention not only in field of natural language processing but is also dominating in the area of recommendation systems to provide effective results which can enhance user online experience. Combining the techniques of deep learning with the architecture of recommendation systems is better adapted to capture the non linear characteristics of user behavior or predicting the next point of interest for the tourist to flash the likable advertisements. Recommendation models are mainly classified as collaborative filtering, content based recommender system and hybrid recommender system as shown in Figure 1 based on the type of data given as input to the system [1].

Background

Plenty of examples can be shown how industry got benefited by applying deep learning to recommender systems. The work of [2] have showed in their work that 80% of movies viewed on Netflix came through recommendations. Authors in [3] have discussed a deep neural network based recommendation algorithm for recommendations of videos on YouTube. The approach in [4] shows an App recommender system for Google Play using deep model. Such successful applications have demonstrated a remarkable improvement over the conventional models of Recommendation systems. Last few years have witnessed an exponential growth in research work based on recommender systems using deep learning in academia because of its increasing popularity in industry.

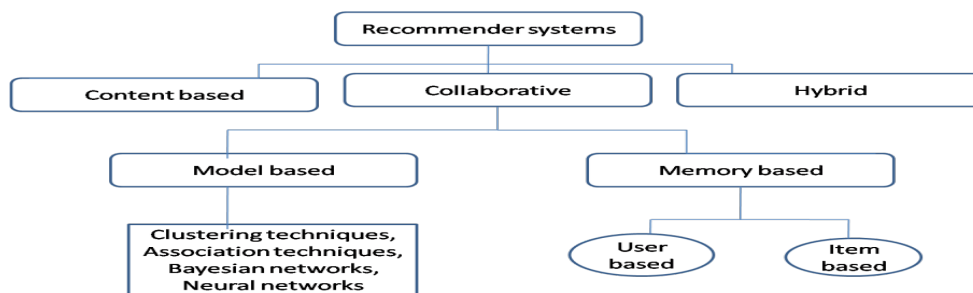


Figure 1: Classification of Recommendation Systems

Recommending the next POI based on LBSN's is very popular in recent times because of which it is gaining lot of attention in the research area. The networking savvy people tend to share not only their check-in but also the experience of the place along with tips or alerts to the other travelers or reader's in general. This data contains lot of information regarding user's choice which can be used to give personalized recommendations for near-by points and can be used by industries to flash the advertisements and promote their product. The authors of [5], [6], and [7] have covered in their research user behavior study, movement pattern analysis, community detection and POI recommendation. Understanding the temporal relations between the series of check-in's is very crucial to give apt recommendation to the users. Authors of [7] in their work used a metric embedding approach to create a personalized model of sequential series of POI's. Authors of [8] have focused on the modeling sequences of venue categories. Their approach uses a mixed hidden Markov model to predict the next likely spot and predict the POI nearest to the neighborhood of the user. Recurrent Neural Networks (RNNs) have been used recently to model sequences because of the effective computational power and advanced architectures [9] [10]. RNN is very useful in language modeling for recursively predicting the next word in the sentence [10].

Discussion

This part presents the analysis of shortlisted papers for conducting the comparative research of deep learning methods for recommending the users with next POI. A set of well known methods like matrix factorization, Word2Vec and RNN are focused in this section.

Novel Matrix Factorization Method: FPMC-LR: Authors of [11] in their work have proposed a matrix factorization method named FPMC-LR to incorporate personalized Markov chains and localized regions. This method makes use of Markov chain to exploit the check-in sequence but also considers user's movement in and around the localized region. Suggesting such personalized POI's based on LBSN's can find variety of applications in giving tips to user's like which restaurant the user can visit to have dinner after movie or which books to buy depending on author's previous taste. Large scale LBSN datasets like Foursquare given by [12] and Gowalla provided by [13] are used to analyze the spatial-temporal properties of LBSN's in this work. The authors have used data of four months from May-2010 to August 2010. The constraint which is used in this dataset is a check-in threshold of 120 times and minimum history of five visits. Table 1 shows the statistics of data derived from Foursquare and Gowalla during the period of May-2010 to August 2010.

	#U	#L	#Check-in	#Avg.check-in
Foursquare	3571	28754	744055	208.36
Gowalla	4510	59355	873071	193.58

Table 1. Statistical Data From Foursquare and Gowalla.

Following steps summarize the working of FPMC-LR:

- A. Successive POI Recommendation:
 - i. First problem to be addressed is to suggest the most suitable recommendation to the user based on the sequence of check-in's and geographical location of the user.
 - ii. Second step attempts to handle the dynamic nature of POI's. It consists of predicting the new location to the user which is not visited before but should be recommended in the consecutive time-stamp.
 - iii. Finally it is important to tap the nature of inter check-in dynamics if any. It is a key factor for finding out relation among successive check-in's in shorter intervals of time. For restaurants or popular eating joints are often visited after watching a movie. This is very synonymous to personalized Markov chain property and can be used to solve the problem of personalized recommendation.
- B. FPMC –LR Model: The model considers only the neighbourhood locations unlike conventional FPMC model [13]. Precisely, complete globe is divided into square grids and then for location, the neighbours will fall in one of the nine adjacent square grids. FPMC-LR produces a transition tensor which is reduced largely for instance around 100 when neighbour is at about 40kms. Therefore FPMC-LR is more time efficient as compared to FPMC. Low rank approximation is an effective means to recover the transition tensor when it is sparse. Recommending top-n new POI's can be modelled as ranking of locations.
- C. Testing: The conducted experiments are tested using the check-in history as training data and last check-in is used as the test data. The crux of recommendation lies in the fact that infrequently visited POI's should be predicted and recommended instead of frequently visited points. The precision and recall values of this experiment are low which justifies the fact that recommending new POI's to the user is more difficult task to do than recommending the frequent ones.

Comparison with FPMC and PTF:

- i. The results of FPMC-LR show that this novel model performs significantly better than FPMC. FPMC-LR gives around 30% and 40% over FPMC for precision and recall respectively. This comparison justifies considering the localized region of the user to remove outliers from the data.
- ii. FPMC-LR remarkably performs better than the FPMC proposed by the work of [14] as well as the Probabilistic Tensor Factorization(PTF)model proposed by [15]. This comparison concludes that Markov chain is vital while giving personalized POI recommendations.

Recurrent Neural Network Model:

The work of [15] have considered temporal relations in between POI categories in tourist paths. Recurrent neural network is used to generate new sequences by training the system over a set of observed paths. The popular Foursquare LBSN dataset is used to gather user's check-in which are ultimately narrowed down to categories of venues and are classified into a set of temporally neighboring activities termed as "paths". For example if the tourist has spent his evening at a city museum then what is the likelihood of him visiting the multi-cuisine restaurant or a karaoke bar which is nearby or what is better after the morning breakfast at a local food joint a historical monument or adventure park. The working of the model can be explained with the help of following steps.

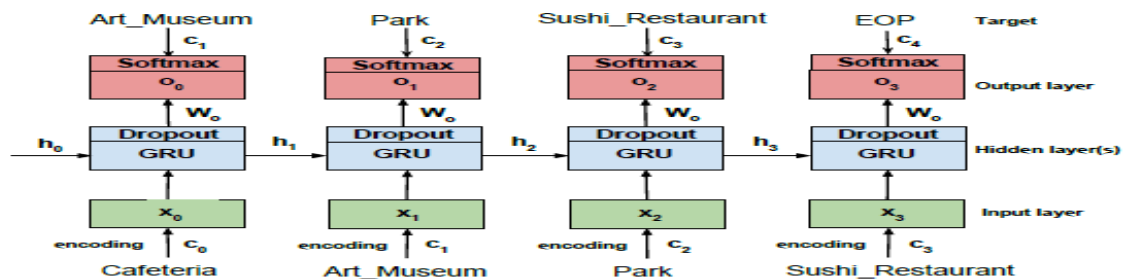


Figure 2: RNN Architecture [15]

- A. Next POI Category Prediction: Machine learning approach is used on temporal sequences of check-in's. The next category of the path is computed in the form of conditional probability applied on all available paths from the sequences of POI categories which yields the next temporal POI category.
- B. RNN Model: This is a specialized RNN which can handle temporal data. The idea of using RNN in predicting POI is very similar to its application in next word prediction. The venue category is fed into the network through encoding in an input vector. This vector is passed to a Gated Recurrent Unit (GRU), which is a gating mechanism that improves the ability of RNN to store longer sequences which have proved to be more effective. The Dropout layer is applied before computing the output which works as a regularization mechanism at training time to randomly cut down on a fraction of neurons called as drop-out rate to avoid co-adapting. Followed by the dropout layer, the output state is defined by a weighted layer whose weights are assigned by a predefined matrix during the training time. Finally, the Softmax layer is used to normalize the output and model the probability distribution of the next category.
- C. Feature Learning: Foursquare Taxonomy is used to define the set of possible categories which classifies the categories using hierarchical ontology. Very popular approach to encode categories is one-hot representation. It is a sparse representation where a binary vector is used whose dimension is equal to the size of the vocabulary and only one component is different from 0 based on the category index. However, this method has a set of shortcomings like wastage of computational resources when some categories are not identified and also fails to identify hierarchical relations among categories. Thus this method makes use of unsupervised feature learning approach called node2vec [16]. Node2vec can retain the structure of the graph even after mapping nodes to vector representations in Euclidean space of fixed dimension.
- D. Testing: It is very crucial to define the metric for evaluating the performance of the model. A very unique metric used in this work is perplexity which is the exponential value of the average negative likelihood of the model. Precisely perplexity can evaluate the surprise element of the model in analysing the test data. Hence, lesser the value of perplexity better is the performance of the model.

Comparison with One-hot Encoding:

Taking into the consideration the results of this model it is evident that combination of RNN and node2vec yields good recommendations as compared to the other contemporary methods. Also, RNN with one-hot encoding gives better results than the bigram model. But when categories are initialized using node2vec encodings the performance of the model is better as compared to standard one-hot encoding.

Word2Vec: Skip-gram and CBOW:

The work of [18] makes use of well-known methods of Word2Vec namely skip-gram and continuous bag of words (CBOW) to give recommendations based on LBSN's. Word embedding i.e., distributed word representations are used to capture linguistic regularities and semantics from large amount of textual information. As in the previously explained models, Foursquare check-in dataset from January-2011 to December-2011 is used for evaluation.

CBOW technique looks around the neighboring words of the target word to predict the target word while skip-gram technique is the exact opposite i.e., it makes use of the target word to predict its neighboring words. Similarity in both the methods is the order of the sequence of words in input does not affect the result. In this work both skip-gram and CBOW is used for recommendation. The approach of the model constitutes of following steps:

- A. Modelling the input data: Word2Vec techniques accept a list of sentences which in turn a list of words that can be used to create an internal repository to hold words and their frequencies together. After this preliminary step the model is trained using input data and dictionary. In this method, along with the list of items, list of users and their past preferences are given as input to Word2Vec techniques. This gives a separate continuous vector representation of item and users. Now this output can be used as feature for making recommendations. The basic idea which drives Word2Vec technique is very synonymous to recommendation which maps past preferences of the users to current choice.
- B. Recommendations using continuous vector representations: The input data is modelled in such a way in the previous step so that the similar vectors lie close to each other and appear as neighbours. Following are the three different recommendation techniques which can be applied over the continuous vector representations:
 - i. K-nearest item (KNI) approach: This approach follows content-based filtering recommendation method. While content-based methods make use of tags that describe users and items KNI makes use of continuous vector representations calculated in modelling step. The cosine similarity between the target user and item vectors are calculated and the top k-items out of the most similar k-items are recommended to the target user.
 - ii. N-nearest users (NN) approach: This approach applies user-based collaborative filtering on continuous vector representations. This approach works in two parts, first the top-n neighbours are found depending on the user vector representations and then the items preferred or bought by top-n neighbours are collected which sums up to the recommendations to be given to the user.
 - iii. N-nearest users and K-nearest items (KIU) approach: As the name suggests this approach combines the ideology of previous two approaches. Initially, top-n neighbours are chosen from user vector representations. Next, top-k items similar to the combination of user and neighbours are shortlisted from the vector representation of the first step.
- C. Testing: The experiments conducted on the test data shows that it is faster to train the model using skip-gram than CBOW. Among the three approaches proposed for recommendation, KNI approach is faster than other techniques followed by KIU. To produce the continuous vector representations, it is analysed that skip-gram techniques give better results than CBOW.

Comparison between KNI, NN and KIU Methods:

The comparative results show that KNI and matrix factorization methods are better than their counterparts NN and KIU. The performance of NN approach is not so promising which shows that recommending the choice of the neighbors is not fully capturing the essence of recommendation.

Conclusion

This article gives a review based on three significant techniques in the domain of deep learning for POI recommendation. The review consists of the analysis of the techniques along with its characteristics and comparison with other stated methods. The novel matrix factorization method FPMC-LR investigates the spatial-temporal properties of LBSN datasets. Although localization constraint plays a vital role in this model for successive POI recommendation in FPMC-LR, it reduces the computation cost and also discards the noisy information which improves the quality of recommendation. The RNN based approach can model and predict the POI categories to recommend a personalized tourist path. This RNN model shows one loophole and that is indefinite interpretation in case of different user clusters where some cases show improvement while in some cases performance drops. Word2Vec techniques although preliminary but the analysis shows that the performance of skip-gram techniques is better than the results achieved through CBOW. Deep learning methods have a huge potential and can definitely enhance the traditional recommender systems. The reviewed work shows that still

there is a scope to use heterogeneous unstructured data samples like visual, audio and video features as an input for the deep learning methods. Feature engineering is in widespread use in industries. But since it is very tedious to manually craft the selected features, deep learning can play a vital role to automate feature engineering.

References

- [1] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 6, pp. 734–749, 2005.
- [2] C. A. Gomez-Uribe and N. Hunt, "The Netflix Recommender System," *ACM Trans. Manag. Inf. Syst.*, vol. 6, no. 4, pp. 1–19, 2015.
- [3] P. Covington, J. Adams, and E. Sargin, "Deep Neural Networks for YouTube Recommendations," *Proc. 10th ACM Conf. Recomm. Syst. - RecSys '16*, pp. 191–198, 2016.
- [4] H.-T. Cheng *et al.*, "Wide & deep learning for recommender systems," *Proc. 1st Work. Deep Learn. Recomm. Syst.*, pp. 7–10, 2016.
- [5] J. Yuan, Y. Zheng, and X. Xie, "Discovering regions of different functions in a city using human mobility and POIs," *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discov. data Min. - KDD '12*, p. 186, 2012.
- [6] Y. Zheng and X. Xie, "Learning travel recommendations from user-generated GPS traces," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 1, pp. 1–29, 2011.
- [7] Z. Cheng, J. Caverlee, K. Y. Kamath, and K. Lee, "Toward Traffic-Driven Location-Based Web Search Categories and Subject Descriptors," pp. 805–814.
- [8] J. Ye, Z. Zhu, and H. Cheng, "What's Your Next Move : User Activity Prediction in Location-based Social Networks," *Sdm*, pp. 171–179, 2013.
- [9] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling," pp. 1–9, 2014.
- [10] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [11] C. Cheng, H. Yang, M. R. Lyu, and I. King, "Where You Like to Go Next : Successive Point-of-Interest Recommendation," pp. 2605–2611, 2012.
- [12] Z. Cheng, J. Caverlee, K. Lee, and D. Z. Sui, "Exploring Millions of Footprints in Location Sharing Services," *Icwsm*, vol. 2010, no. Cholera, pp. 81–88, 2011.
- [13] C. Cheng, H. Yang, I. King, and M. R. Lyu, "Fused Matrix Factorization with Geographical and Social Influence in Location-Based Social Networks," pp. 17–23.
- [14] S. Rendle, C. Freudenthaler, and L. Schmidt-Thieme, "Factorizing personalized Markov chains for next-basket recommendation," *Proc. 19th Int. Conf. World wide web - WWW '10*, p. 811, 2010.
- [15] L. Xiong, X. Chen, T.-K. Huang, J. Schneider, and J. G. Carbonell, "Temporal Collaborative Filtering with Bayesian Probabilistic Tensor Factorization," *Proc. 2010 SIAM Int. Conf. Data Min.*, pp. 211–222, 2010.
- [16] E. Palumbo, G. Rizzo, and E. Baralis, "Predicting Your Next Stop-over from Location-based Social Network Data with Recurrent Neural Networks," pp. 1–8, 2017.
- [17] A. Grover and J. Leskovec, "Node2Vec," *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. - KDD '16*, pp. 855–864, 2016.
- [18] M. G. Ozsoy, "From Word Embeddings to Item Recommendation," 2016.