Toward a Context-aware Serendipitous Recommendation System

Changhun Lee, Gyumin Lee, and Chiehyeon Lim*

School of Management Engineering
Ulsan National Institute of Science and Technology
Ulsan, 44919, Republic of Korea

Abstract

Recommendation system development has been an important domain in the industrial and academic fields for the past two decades. Recently, the importance of developing a context-aware serendipitous recommendation system has emerged. As such, we investigate the latent features of items that may be recognized by the users of such a system. We assume that users will move from one item to another through the latent features reflected in the sequence of items. Our work specifically focuses on the process of predicting the sequential and changing taste of users. We show the existence of latent features by presenting a topic map and suggest a context-aware serendipitous recommendation system.

1. Introduction

Recommendation system development has been an important domain in the industrial and academic fields for the past two decades. The primary goal of this system is to provide users with personalized items based on past records to improve their satisfaction. However, as techniques improve and research on recommendation systems increases, researchers are facing the question of how well a user is satisfied with the recommendation (i.e., the value of the recommendation). For example, the importance of developing a "serendipitous" recommendation system has emerged as part of improving the value of the recommendation system.

Thus far, accuracy is considered a representative measure for estimating the value of a recommendation system. Accuracy indicates the probability that the user will appreciate the item recommended [2]. Although using this measure sounds simple and logical, a few researchers argue that other things should be taken into account [5; 6; 9]. As a result, the necessity of measuring "diversity" and the concept of serendipitous recommendation has emerged. Diversity is a literal indicator of how diverse items are included in the recommended set and is therefore inevitably associated with the concept of serendipitous recommendation. However, diversity alone does not secure a serendipitous recommendation, because serendipity means both unintended and useful discovery (i.e., unexpectedly satisfactory discovery) [5; 11]. In other words, a recommendation system built specifically to increase diversity will recommend novel items regardless of the user's satisfaction. For example, given a recommendation system that recommends unfamiliar movies to the user on purpose, the user will probably be dissatisfied with the recommended movies because his/her taste is sacrificed for diversity. On the other hand, for a system that recommends familiar movies only, it is likely that the user's taste will be over-reflected such that the system will recommend accurate but obvious movies (i.e., movies that the user would have discovered by himself/herself). This is not a serendipitous recommendation system because the recommended movies are not unexpected. This situation is a well-known trade-off relation between diversity and accuracy. Hence, many studies focused on increasing the diversity while minimizing the accuracy loss for a serendipitous recommendation [7; 12], and our work likewise is conducted in a similar way.

Seeing that serendipity alleviates the trade-off between diversity and accuracy, we need to define the concept of diversity and figure out how to measure it. To define and measure diversity, we paid attention to the latent features of items. We assumed that people would move from one item to another through a latent feature reflected in a sequence of items. To sum up, we considered that there are latent features that link each item and the variety of the topics measures the diversity.

* Corresponding author: Tel.: (052) 217-31120; E-mail: chlim@unist.ac.kr

Specifically, in this study, we developed a movie recommendation system with user rating data. Our data included who the users were and how and when they rated movies with a certain score. The ratings were recorded chronologically and could thus be regarded as a list of consecutive movies that reflect the latent features. Our work specifically focused on the impact of the latent features by treating a recommendation as a process of predicting the sequential and changing taste of users. We indirectly show the existence of latent features by presenting a topic map and suggest a context-aware serendipitous recommendation system.

This paper is organized as follows. First, Section 2 addresses the main idea on which our recommendation system is built based on a concept from cognitive psychology field. Then, Section 3 presents a detailed description of the algorithm used in our work. Sections 4 and 5 explain the metrics and data, respectively. Finally, Sections 6 and 7 cover the evaluation and conclusion respectively.

2. CONTEXT : SPREADING ACTIVATION MODEL

In cognitive psychology, one of the most famous theories regarding human semantic processing is spreading-activation theory. This theory is based on Quillian's theory of semantic memory [3]. Qullian viewed memory search as an activation spreading from the concept nodes through a semantic network [8]. For example, when people are asked to state everything about machines (stimuli), they start off with clear facts (reaction), such as "it is manmade," "it has moving parts," and so on. Soon, however, they begin to give less clear facts, such as "a typewriter is machine." This stimuli–reaction mechanism is represented by a spreading-activation model, which consists of nodes and links. Each node corresponds to the concept that people can recall as a reaction against the stimuli, and the relational links indicate how strongly the concepts are related. Nodes preferentially activate the peripheral nodes, which are strongly related, and the activation spreads in a way that the activated nodes activate other linked nodes, and so on.

Motivated by the spreading-activation model, we assumed that people semantically link movies on the basis of latent features. For instance, a user logs into a rating website (e.g., IMDB or MovieLens) to rate a movie he/she just watched. After the user rates the first movie (stimuli), the movies he/she thinks are relevant (reaction) will pop up in his/her head and he/she will probably rate one of them. This process will continue until a user log out. What makes the user rate consecutively from one movie to the next is the latent features. [1] We call this concept "context."

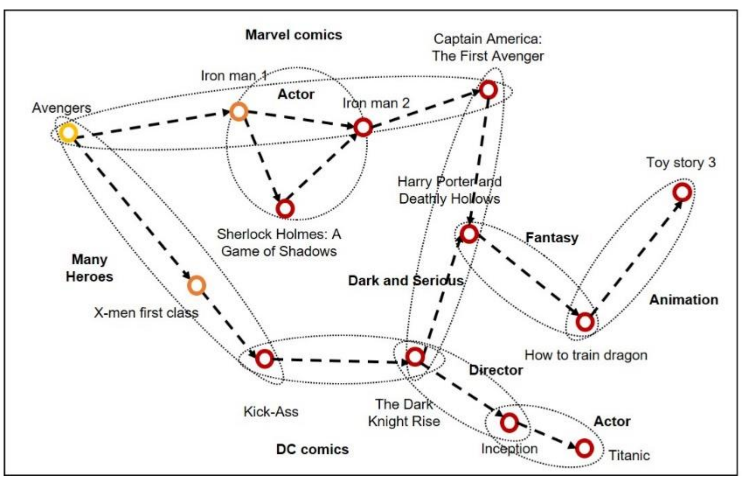


Figure 1: Example of Movie Semantic Network

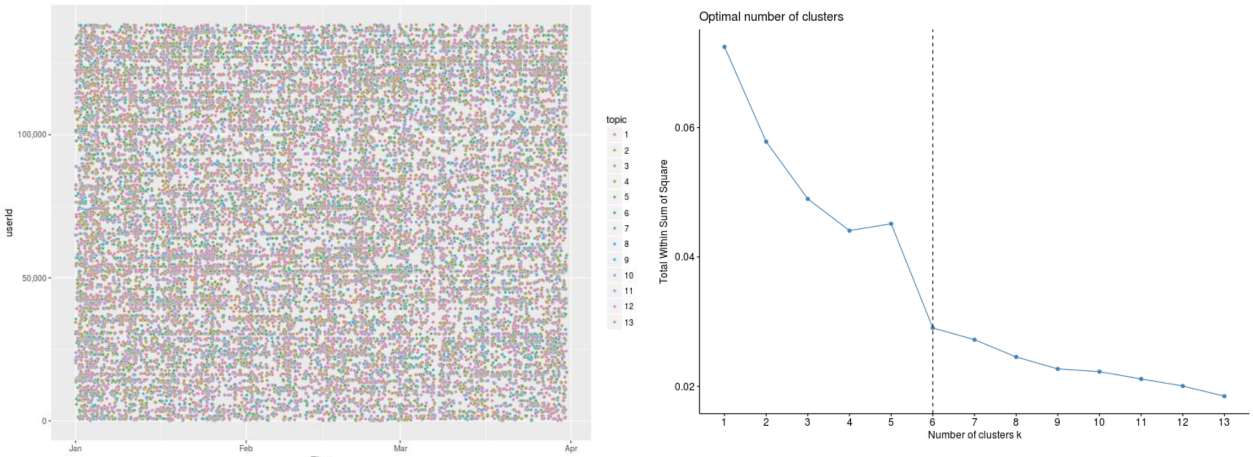
3. ALGORITHM

Two popular techniques of information retrieval are used to extract and train the latent features, namely, Latent Dirichlet Allocation (LDA) and Word2Vector (W2V). LDA provides a global picture of latent features (e.g., the number and type of latent features in data) by allocating every movie into every topic, while W2V vectorizes each movie with local information related to its latent feature.

[1] Those features could be the actor, director, genre, series (e.g., Marvel comics), animation, japanimation, atmosphere and so on

3.1. LATENT DIRICHLET ALLOCATION

LDA is a topic modeling technique [4]. Basically, it was developed to extract latent features (i.e., topics) from a corpus using the overall structure of the words in documents. [1] This is why we used LDA. We wanted to know which and how many topics exist in the overall data. As a result, a total of 13 topics were derived. Figure 2a shows the change in topics per user by time. Specifically, it visualizes the route by which a user passes through his/her own semantic network, such as shown in Figure 1.



(a) Topic Map (b) Finding k

Figure 2: Topic Map and Finding *k*

LDA is a soft clustering technique. In LDA, each movie is non-exclusively allocated to each topic with a specific probability, but what we want is to exclusively allocate each movie to each topic. Therefore, additional work was carried out to transform the result of LDA into a hard clustering format. First, we made a movie–topic matrix. The values in matrix are the probabilities of being allocated to each topic per movie. Then we applied k-means clustering to exclusively allocate each movie to each topic. The proper number of k was determined by the elbow method, as shown in Figure 2b. In our case, $k = 6$.

3.2. MOVIE2VECTOR

Our recommendation algorithm, Movie2Vector, is an application of W2V, which is a famous word embedding technique proposed in [10]. When a bundle of corpus is given, W2V technique trains local information per word by a moving training window. This technique is based on the concept that the meaning of a word is locally defined in a relationship with the neighborhood words. We assumed that the movie sequence per user has a similar property with the word sequence in the document and therefore applied W2V in our recommendation system.

4. METRICS

We used two metrics in our work, namely, the accuracy and diversity defined in this section. Before proceeding to the metrics, let us briefly explain our notations. $Q_u = \{Q_1, ..., Q_n\}$ notates a set of query movies per user and is used as input. $R_u = \{R_1, ..., R_n\}$ denotes a set of recommended movies, that is, the output. Test set per user $T_u = \{T_1, ..., T_n\}$ is the sequence of the movies after Q_u, where Tu indicates a set of target movies following a query set. TP is a set of topics. A mapping function $M(*)$ is also utilized to map a set of movies into a set of topics to which each movie is allocated.

4.1. ACCURACY

[1] LDA allocates the words into every topic based on the Dirichlet distribution.

Hit rate indicates how many items in T_u are hit. In other words, it shows how many items in R_u matches with items in T_u. This is a simple and popular metric to measure accuracy and was therefore employed in this study.

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{n}(\boldsymbol{T_u} \cap \boldsymbol{R_u})}{\boldsymbol{n}(\boldsymbol{T_u})} \tag{1}$$

4.2. DIVERSITY

We defined our own diversity metric. We assumed that the diversity originates from the diverse latent features and thus measured the diversity on basis of a variety of topics. The metric is defined as a ratio of new topics that appeared in R_u over the total number of T.

$$\boldsymbol{Diversity} = \frac{\boldsymbol{n}(\boldsymbol{M}(\boldsymbol{R_u})) - \boldsymbol{n}(\boldsymbol{M}(\boldsymbol{Q_u}) \cap \boldsymbol{M}(\boldsymbol{R_u}))}{\boldsymbol{n}(\boldsymbol{TP})} \tag{2}$$

5. DATA

Choosing an appropriate dataset is crucial in evaluating an algorithm. Our work uses the rating sequence of each user, meaning that we need user rating, timestamp, and the identifier of the corresponding movies. The confidence of dataset needs to be considered as well. We assumed that the more referred the dataset was, the more confident it would be. The two criteria (appropriateness and confidence) led us to use the MovieLens dataset. MovieLens is a website that provides a movie rating service[1] run by GroupLens, a research laboratory at the University of Minnesota. It has offered various types of datasets since 1998, and the datasets are basically given by size, from 100 K to 20 M, each of which contains genome scores, genome tag links, movies, ratings, and tag files. The size and file construction of the MovieLens dataset makes it worthy of use and popular in the recommendation system research field, such that there is even a paper which covers the history of the dataset [1]. As of May 28, 2018, when you search "movielens" in GoogleScholar, you can retrieve about 14 K results. This figure means a fairly large number of articles are referring to the dataset, which proves its confidence. Among the files in each dataset, the ratings file includes information about users' preference. This file takes the form of ⟨user id, movie id, rating, timestamp⟩. Overall, the MovieLens dataset achieves both appropriateness and confidence and is, therefore, the best dataset for applying our algorithm. Our dataset from MovieLens includes 20 M ratings and 46.5 K tag applications across 27 K movies by 13.8 K users between January 09, 1995 and March 31, 2015. It was first released on April 2015 and most recently updated on October 2016. According to the data provider, the dataset was constructed by random extraction only for users who had rated at least 20 movies.

6. EVALUATION

We argue that latent features exist and move users from one movie to next, which represents the change in context. Although the latent features are invisible, we considered them definitely hidden behind the sequential records of the rated movies. Our recommendation system was evaluated if it reflected this perspective.

The result of recommendation is plotted in Figure 3. We compared our algorithm (W2V) to a user-based collaborative filtering (CF) and the random recommendation (RD) algorithm. Every point in Figure 3 indicates a result of a recommendation per user. We ran each algorithm on randomly selected 300 users. If we carried out random sampling once per user, the result of recommendations can do over-fitting to the specific query set of the users. Thus, a cross-validation was performed. The number of validations was set to 20, and it described that every point in Figure 3 represents the per user mean diversity and mean accuracy of the recommendations derived from 20 times of random sampled queries. Figure 3 and Table 1 clearly show that W2V performs better than CF and RD on the accuracy perspective. This finding suggests that context-awareness is a noteworthy factor in recommending an item.

Additionally, the way CF observations are distributed shows a negative slope, which explains the trade-off relation between accuracy and diversity. On the contrary, W2V displays a relatively flat distribution. The weakening trade-off in W2V once again supports our argument that context-awareness is important in serendipitous recommendation.

[1] https://movielens.org/

However, our algorithm is limited in terms of the average diversity. This work is an ongoing effort, and the present conference paper shows the in-progress result.

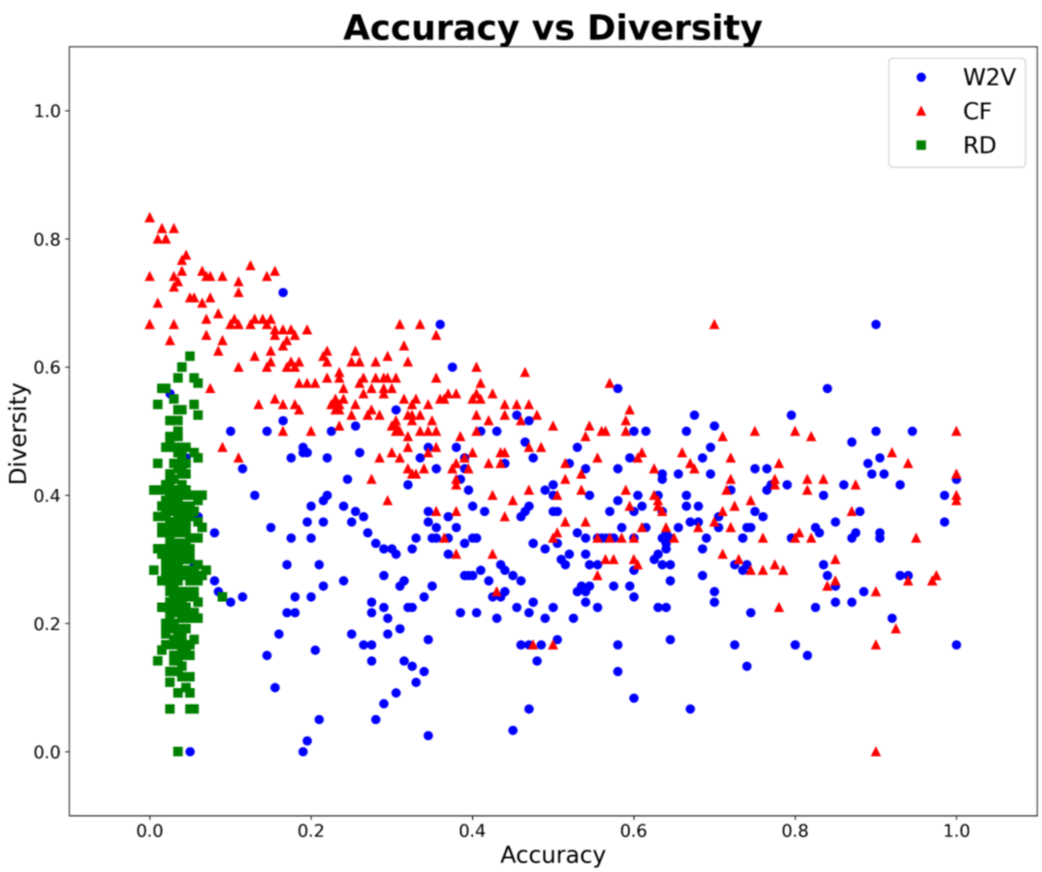


Figure 3: Accuracy vs. Diversity

Table 1: Performance Comparison

Algorithm	**Accuracy**	**Diversity**
W2V	0.499	0.324
CF	0.397	0.510
RD	0.035	0.314

7. CONCLUDING REMARKS

One cognitive psychology theory, the so-called spreading-activation theory explaining about human semantic processing, motivated us to develop the proposed recommendation system. We assumed that the cognitive factor affects the rating behavior of users and must therefore be considered in the recommendation system designed for her/him. In the proposed system, the semantic links in spreading-activation theory are mapped into the latent features (i.e., topics), and we termed them as context. We applied the LDA and W2V for our algorithm to be context-aware. Then, we showed that our context-aware recommendation system performs better on the accuracy perspective, which proved the importance of context-awareness.

Although our algorithm alleviates the traditional trade-off problem in recommendation systems, thus far, our interim result only seems to be context-aware-accurate not serendipitous. Nonetheless, we think the resulting non-pattern (circle dots in Figure 3) at least shows a possibility of being serendipitous. We are continuing to improve our algorithm for it to be serendipitous by increasing diversity while minimizing accuracy loss.

We conclude this paper by emphasizing its relevance to the service science. Nowadays, recommendation systems have become indispensable elements in many service systems, from item recommendation in e-commerce services to friend recommendation in social network services. The development of high-performance recommendation algorithms will soon become the core competitiveness of many service firms.

REFERENCES

[1] *F. Maxwell Harper and Joseph A. Konstan, "The MovieLens Datasets: History and Context", ACM Trans. Interact. Intell. Syst. 5, 4, Article 19, pp.1-19, 2015.*

[2] *Cai-Nicolas Ziegler et al. "Improving recommendation lists through topic diversification", Proceeding WWW '05 Proceedings of the 14th international conference on World Wide Web, pp.22-32, 2005.*

[3] *Collins, A. M., & Loftus, E. F. : "A Spreading-Activation Theory of Semantic Processing", Psychological Review, Vol.82, No.6, pp.407-428, 1975.*

[4] *David M. Blei et al. : "Latent Dirichlet Allocation", Journal of Machine Learning Research, Vol. 3, pp.993-1022, 2003.*

[5] *Jonathan L. Herlocker et al. : "Evaluating Collaborative Filtering Recommender Systems", ACM Transaction on Information System (TOIS), Vol. 22, Issue 1, pp.5-53, 2004.*

[6] *Mouzhi Ge et al., "Beyond accuracy: evaluating recommender systems by coverage and serendipity", RecSys '10 Proceedings of the fourth ACM conference on Recommender systems, pp.257-260, 2010.*

[7] *Neal Lathia et al., "Temporal Diversity in Recommender Systems", SIGIR '10 Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval, pp.210-217, 2010.*

[8] *M. Ross Quillian. : "Word Concepts: A Theory and Simulation of Some Basic Semantic Capabilities", Behavioral Science, Vol. 12, 1967.*

[9] *Sean M. McNee. : "Being Accurate Is Not Enough: How accuracy metrics have hurt recommender systems", Proceeding CHI EA '06 CHI '06 Extended Abstracts on Human Factors in Computing Systems, pp.1097-1101, 2006.*

[10] *Tomas Mikolov et al. : "Efficient Estimation of Word Representations in Vector Space", arXiv preprint arXiv:1301.3781., 2013.*

[11] *Elaine G. Toms., "Serendipitous Information Retrieval", In DELOS workshop: information seeking, searching and querying in digital libraries, pp.17-20, 2000.*

[12] *Zhang, Y. C. et al., "Auralist: introducing serendipity into music recommendation", In Proceedings of the fifth ACM international conference on Web search and data mining, pp.13-22, 2012.*