Journal of Theoretical and Applied Information Technology
15th September 2020. Vol.98. No 17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

IMPROVING RECOMMENDATION ACCURACY AND DIVERSITY USING NETWORK EMBEDDING METHOD

[1]NADIA BUFARDI, [2]HAYTAM HMAMI, [3]ABDELHADI FENNAN, [4]ABDELOUAHID LYHYAOUI

[1,4]Laboratory of Innovative Technologies, National School of Applied Sciences, Tangier, Morocco

[2,3]Computing, Systems and Telecommunications Laboratory, Faculty of Science and Technology, Tangier, Morocco

E-mail: [1]boufardi.nadia@etu.uae.ac.ma, [2]hmami.h@gmail.com, [3]afennan@gmail.com, [4]a.lyhyaoui@uae.ac.ma

ABSTRACT

In recent years, diversity in recommender systems have become increasingly an essential dimension for evaluating the effectiveness of recommendations. However, many existing recommendation techniques are challenged by information overload with the widespread use of recommender systems in many real-world applications. In this paper, we propose a new diversified recommendation approach, namely DRN2V, based on rich constructed graphs and Network Embedding technology. Specifically, we construct a knowledge graph of two sub-graphs, the User-Item subgraph that represents the interactions between users and items and the Item-Category subgraph which uses the item categorization to enrich the network structure. Afterwards, we use Node2vec algorithm to capture the complex latent relationships between users and items from the constructed knowledge graph. Moreover, to propose personalized and relevant predictions for each user, a new formula was proposed based on category coverage and users' preferences for categories. The experimental results demonstrate the significant outperforms of our approach over several embedding-based methods and recommendation algorithms including both traditional and diversity-oriented algorithms) regarding accuracy and diversity.

Keywords: *Recommender system, Accuracy, Diversity, Network embedding, Node2vec*

1. INTRODUCTION

Since the appearance of first research papers on recommender systems in the mid-1990s [1]–[4], accurately predict the user preferences were directly or implicitly the main motivation for recommender systems. However, a broader perspective on the usefulness of recommendation, including but going beyond accuracy, began to appear in the literature in the early 2000s [5], [6], realizing the importance of diversity among other properties as an essential ingredient influencing user satisfaction [7]–[9].

Diversification in recommender systems describes the ability of the system to propose items that users would not discover themselves [9]. It has been firstly defined as being the opposite of similarity such that the recommended items are dissimilar between them [10], [11]. While more recent research [12], [13] specified that diversity mainly represents a variety of elements in the lists of recommendations. Besides, diversity can be viewed at two levels: individual or aggregate. Individual diversity is specific to each user, and it attempts to increase dissimilarity between elements of the recommended list for the target user [14]–[17]. In contrast, the aggregate diversity corresponds to the diversity of recommendations at the system level [18]–[22], by measuring for example the total number of distinct elements among the lists recommended to all users [23].

When the diversity issue was raised in recommender systems, several researches have been proposed to overcome this issue [14], [17], [20], [24]–[26]. Generally, most of these approaches described in the literature consist in using a heuristic strategy that reclassifies all recommendations on two stages: (1) application of existing recommender system to obtain a set of accurate candidate items, and then (2) selecting the Top-N items from the candidates by maximizing a specific diversification criterion [14], [26]–[29]. However, these re-ranking techniques depend directly on the generated list of candidate items, which must be already diversified. Another class of approaches circumvents this problem, diversity modeling approaches [30], which aims to improve the recommendation algorithms by integrating diversification analysis before reclassification phase to generate recommendations, so that precision and diversity are simultaneously taken into account [31]–[36].

However, with the widespread use of recommender systems, the information networks representing user and item content as well as their interactions has become increasingly high-dimensional and sparse networks [37], which makes network analysis a complex task and computationally expensive for many existing techniques. Therefore, Network Embedding (NE) algorithms have been proposed to overcome these challenges by embedding network vertices into a low-dimensional vector space while preserving network topology structure and vertex content [38].

In this work, we address the diversity issue of recommended items along with the sparsity of recommender system networks. Hence, we are interested in recommending a variety of items in terms of categories or genres by harnessing the power of NE algorithms applied in the recommender system network in order to increase diversity of the recommended list while maintaining accuracy. Toward this end, we propose a new Diversified Recommendation approach based on Node2vec method called DRN2V. In DRN2V approach, Node2vec (N2V) learns low-dimensional vector representations for different nodes of a constructed knowledge graph. Afterwards, a new formula was proposed based on category preferences and category coverage in the recommended list to propose personalized and relevant predictions for each user.

The rest of the paper is structured as follows. Section 2 provides a brief survey of related work in three sub-section: first, we describe some classical NE algorithms, then we present some researches applying NE technology in recommender systems, and lastly, some diversified recommender systems based on NE algorithms are discussed. Section 3 introduces our proposed approach. Experimental results are reported in Section 4, and we conclude with a discussion of the proposed work and highlight some future research directions in section 5.

2. RELATED WORK

2.1 Network Embedding (NE)

In recommender systems, the user-item interactions can be represented as knowledge graphs that can be easily enriched by different resources available in the web to further enhance the performance. At this stage, the crucial point is to model user-item interactions from these rich knowledge graphs in order to make relevant recommendations. Accordingly, many NE models has been shown their effectiveness to properly represent a knowledge graph by learning low-dimensional vector representations for nodes or edges in the graph. Inspired by word2vec [39], which adapt neural language models to learn word embedding based on a trained neural network, two popular models for feature representations in network have been proposed: DeepWalk [40] and Node2vec [41]. In [40], Perozzi et al. proposed DeepWalk approach that transform each node in a network to a vector of latent representations. The authors process a neural language model on sequences generated by using random walks on a graph and forms network embedding's with skip-gram model optimized using Hierarchical Softmax. Although DeepWalk has shown its efficiency and scalability for analyzing massive data, the algorithm suffers from an inefficient random walk strategy and its incapability to use negative sampling. Therefore, Grover and Leskovec [41] proposed an improvement over DeepWalk named Node2vec. The algorithm adopts a more sophisticated random walk strategy and uses negative sampling to enhance network feature learning. It learns low-dimensional representations for nodes with a neighborhood sampling strategy, i.e., biased random walk, which smoothly interpolates between two extreme sampling strategies: Breadth-first Sampling (BFS) and Depth-first Sampling (DFS).

2.2 Embedding-based Recommender Systems

Along with the development of NE, many embedding-based recommenders' systems were proposed specifically using Node2Vec, to take advantage of the learned representations for achieving better performance [42]. For instance, Palumbo et al. [43] proposed a recommendation approach named Entity2rec that measures user-item relation to recommend items. As input, the approach uses a knowledge graph constructed from users comments and items properties extracted from the Linked Open Data. The authors learn via node2vec the vector representations of nodes by taking into account one property at a time. Then, they combine scores computed from the latent representations of relationships between users and items specific to a property into a global relationship score using a supervised learning approach. After that, Palumbo et

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

al. [44] proposed an improvement of Entity2rec by allowing to generate one knowledge graph embedding's for all properties. They applied node2vec algorithm on a knowledge graph that includes users, items and users' feedback on items as well as other item-related properties (e.g. director, leading actor). Next, the recommendations are generated based on the ranking of a calculated cosine similarity between the users and items vectors. Meanwhile, the authors in [45] have been working on models to reduce the size of the graph used by the NE algorithm. They proposed three methods: one removes the edges which have a weight lower than a threshold along with the nodes which do not have edge, the second method orders the edges by their weights and leaves only the Top-N edges following a threshold and the third, calculates the number N of edges to keep according to a threshold and nodes degrees. Otherwise, Liang et al. [46] proposed new similarity measure to enhance the basic collaborative filtering algorithm through a latent representations of nodes generated using Node2Vec along with a social trust network. Recently, Chen et al. [47] proposed a spectral clustering method based on Node2Vec algorithm to make recommendations. So, after generating the representation vectors using graphs that connect users to each other and items based on the one-mode projection, they cluster users and items based on the closest dynamic neighbor algorithm (DNN) and automatically cluster number determination algorithm (ADCN).

2.3 Diversity in Recommender Systems

In the literature, several researchers have proposed adding diversity in recommendation process to increase user satisfaction. The first research was mainly based on a re-ranking procedure that tries to re-rank a list of candidate items generated by an existing recommender system using some diversity scores [11], [48], [49]. Other researchers proposed to diversify recommended list by selecting items from several clusters [50]–[52]. As in [53] and [33], they partitioned user's profile into clusters of items or sub-profiles and recommend items by treating each sub-profile as being an independent user profile. Furthermore, recent research incorporates a diversity parameter in the learning process [25], [31], [32], [35], [54]. In [55], the authors added a diversity parameter in constrained Probabilistic Latent Semantic Analysis (PLSA) to optimize accuracy and diversity simultaneously. Moreover, the authors in [14] optimized the set of recommended items through structural SVM (Support Vector Machine) with a loss function combining diversity and accuracy.

Besides, there have been a few works on strategies that take advantage of the embedding network methods to enhance diversity. For instance, Pouli et al. [56], one of the first researchers to takes into account diversity using embedding networks. The authors have proposed a recommender system based on hyperbolically embedded social networks to improve the diversity of recommended items by following their most similar neighbors' actions. They start building user-user follower networks using some follower's similarity measure. Then, using the generated networks, the prediction is calculated by selecting the taste of the most similar neighbors to the target user according to their preferences. Furthermore, to ensure forwarding relevant item recommendations to the most interested users, two phases were presented: (1) the authors greedily embeds the minimum spanning tree constructed from the original graph into the hyperbolic disk. (2) a context-based routing algorithm was proposed based on relevance metric that incorporates a context similarity measure and a network distance. Whereas, J.Lin et al. [57] have proposed a recommendation approach to enhance the recommendation diversity for the billion-scale recommendation. The proposed approach is a diversity and dynamics-aware sampling method that is processed in two phases. Initially, a negative sampling method was used to sample some negative edges. Next, items are reduced based on the diversity of user interest (select items of more than one cluster) and the dynamics of user interest (focus more on recent clicked items). Moreover, to preserve the local structures between users and items, a loss function is defined as a sum of empirical distributions on the original network and the probability specified by the low-dimensional representations reconstructed on the embedding space.

Unlike approaches described above, we are interested in our approach in two aspects. One is using the item categorization information by taking into account user preferences for certain categories and trying to diversify its taste without falling into redundancy. The second is respecting the level of diversity of categories for each user to make a personalized diversified recommendation.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3. PROPOSED APPROACH

In order to recommend accurate and diverse items, we propose a new Diversified Recommendation approach based on Node2vec (DRN2V) that is composed of the following three steps, as illustrated in Figure 1:

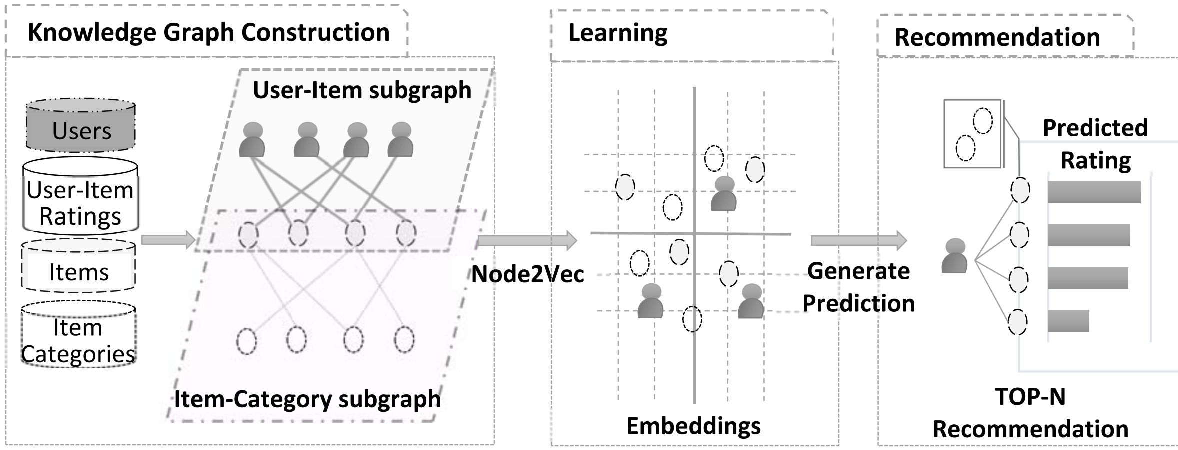


Figure 1. Overview of the proposed DRN2V model

(1) **Knowledge Graph Construction**. Construct knowledge graph based on two relations. The first designates the relationship between users and items. And the second, consists of weighting the relationship between items and their categories.

(2) **Learning Embedding**. Apply the Node2vec algorithm to generate users, items and categories embedding from the entire knowledge graph.

(3) **Recommendation**. Use a new formula based on category preference and category redundancy information to calculate the predicted rating for unseen items to recommend Top–N accurate and diverse items to users.

3.1 Recommendation Problem

Before presenting our proposed approach, we briefly introduce the notation used throughout this paper. Three entities typically identify the data sources considered by recommender systems: a community of users $u \in U$, a list of items $i \in I$ and transactions, i.e. Relationships between users and items which reflect the interest intensity given by the user to an item either implicitly, such as purchase records or clicks, or explicitly, on a numerical scale noted by $r(u,i)$. In addition to this, we assume that a categorization C of each item in I is available, where the relation between items and categories $c \in C$ is weighted by a binary value: if an item $i \in I$ belongs to a category c, then the weight takes the value 1, and 0 otherwise. The recommendation task can be defined as a Top-N recommendation problem. Specifically, for a given user, the overall goal is to find a top N list of ordered candidate items that maximize the estimated scores produced by the recommender system for each candidate item $\hat{r}(u,i)$:

$$LN(u) = \{i_1, i_2, i_3, \ldots, i_N\} \quad (1)$$

where $\hat{r}(u,ij) > \hat{r}(u,ij+1)$ for any $j = 1\ldots(N-1)$.

3.2 Knowledge Graph Construction

The construction of the knowledge graph required two sub-graphs; first, the User-Item subgraph, which describes the users' interests for seen items, second, the Item-Category subgraph, which allows to present item-category relationship by respecting users' preferences.

3.2.1 User-Item subgraph

The User-Item subgraph is defined as $G_{ui} = (U \cup I, E)$, where U is a set of users, I is a set of items and E denotes the set of edges formulated as follows:

$$E = \{r(u,i) : u, i \in |U| \times |I|\} \quad (2)$$

An edge could exist between a user node and item node if the user has expressed his interest in the corresponding item with a weight e. Accordingly, every edge has a weight $e = r(u,i)$ where $u \in U$, $i \in I$ and $r(u,i)$ is the rating given by user u to item i.

3.2.2 Item-Category subgraph

As already mentioned above, the relationship between categories and items is defined by default as a membership relationship. In this work, we aim to take advantage of user preferences information for each category in order to enrich the relations between items and categories. Hence, the Item-Category subgraph G_{ic} is presented as a set of triples (i, s, c) where $i \in I$, $c \in C$ and $s \in S$, and S is a set of relations scores among items and categories. We define the relation $s(i,c)$ between an item i and a category c as follows:

$$s(i,c) = \frac{1}{|U_i|} \sum_{u \in U_i} \frac{\sum_{j \in I_{u,c}} r(u,j)}{|I_{u,c}|} \quad (3)$$

where $U_i = \{u \in U \mid r(u,i) \neq 0\}$ is the set of all users who rated item i and $I_{u,c}$ is the set of items rated by user u and belongs to category c:

$$I_{u,c} = \left\{ j \in I \;\middle|\; \begin{matrix} j\ belongs\ to\ c \\ r(u,j) \neq 0 \end{matrix} \right\} \quad (4)$$

3.3 Learning Embedding Using Node2vec

Though the constructed knowledge graph is enriched with different sub-graphs mentioned above, it is difficult to properly capture network features that best represent the constructed knowledge graph. Therefore, we adopt the Node2vec technique as a NE method to alleviate the data sparsity problem, which has shown its effectiveness and stability due to its flexible neighborhood sampling strategy.

Let $G = (V, E)$ be the constructed knowledge graph encompassing users U, items I and categories C as nodes and E as edges (e.g. ratings or calculated scores) between nodes, the Node2vec algorithm will map each node $v \in V$ to feature vector $f(v)$ using a mapping function $f : V \rightarrow R^d$ with d is the dimension of latent space. So as to learn feature vectors, a set of neighbor nodes $N_s(v) \in V$ of node v is generated by exploring neighborhoods through flexible biased random walk procedure, e.g. it combines micro-view of the neighborhood using breadth-first sampling (BFS) and macro-view of the neighborhood using depth-first sampling (DFS). The trade-off between the two views is achieved by using hyper-parameters p (return parameter that controls the likelihood of backing to the previous node in a walk) and q (in-out parameter that controls the probability of moving outwards (DFS) or inwards (BFS)) to guiding the walk.

Following the skip-gram architecture, the objective function is then optimized by maximizing the occurrence probability of neighbor nodes $N_s(v)$ conditioned on its feature representation given by the mapping function f:

$$\max_f \sum_{v \in V} logPr(N_s(v)|f(v)) \quad (5)$$

3.4 Recommendation

In order to provide an effective item ranking procedure that improves accuracy and diversity at the same time, we propose a new formula that enhances the coverage of different categories while minimizing their redundancy. We consider two proprieties: (1) Category Preference, that measures user preferences for a category using the generated embedding vectors of users and categories. (2) Category Redundancy, which helps ensure that categories covered by the recommended list are not over-represented.

$$\hat{r}(u,i) = \sum_{c \in C_i} pref_C(u,c) \times red_C(u,c) \quad (6)$$

where C_i is a set of categories to which the item i belongs to.

Initially, to measure the two proprieties, we started by initializing a pre-recommended list using the candidate items sorted in descending order according to the number of categories of each candidate item.

Afterward, to define the Category Preference ($pref_C$), we assume that more the category is less preferred by the user, more the presence of this category in the recommended list will allow to make recommendations that deviate from what the user generally prefers. So we used the opposite of the distance between a user u and category c embedded vectors:

$$pref_C(u,c) = 1 - dist(f(u), f(c)) \quad (7)$$

where $dist(f(u), f(c))$ is defined based on the cosine similarity measure:

$$dist(f(u), f(c)) = cos(f(u), f(c)) = \frac{f(u), f(c)}{\|f(u)\| \times \|f(c)\|} \quad (8)$$

In addition, we model the Category Redundancy (red_C) as an exponential decay function that decreases rapidly as the occurrences of category increases, i.e. when a category c appears more than once in a recommendation list, it can be potentially redundant, then the prediction given to the item that belongs to category c must be low. The red_C is calculated as follows:

$$red_C(u, c)@k = \lambda^{occu(c)@k} \quad (9)$$

where the base of the exponentiation is represented by a parameter $\lambda \in [0,1]$ and the exponent $occu(c)@k$ is the number of occurrences of a category c in the pre-recommended list at k-th position.

Time Complexity. The basic Node2vec method holds a time complexity of $O(|V|^2)$. In particular, the method generates $|V| \times r$ walks each of length k where V is set of nodes, and r is the number of walks, then it calculates the node embeddings. Therefore, the DRN2V generating embeddings phase has a time complexity of $O(|V|^2)$, as we construct it by applying Node2vec method. Besides, for prediction phase, according to n as the number of users, m as the average number of items seen by each user and l as the average number of items categories, we note that m and $l << n$, so the rating prediction phase has complexity $\sim= O(n)$.

4. EXPERIMENTS

In this section, we discuss experiment protocols and evaluation results to validate the effectiveness of our proposed approach.

4.1 Datasets

Since we have defined diversity based on item categorization information, the datasets used for evaluation must contain this information. Thus, we conduct our experiment on real-world movies datasets that categorize movies in different genres or categories, e.g. the movie "Get Carter (2000)" belongs to genres Action, Drama, and Thriller. We used three datasets: MovieLens 100K (ML-100K) and MovieLens 1M (ML-1M) datasets, [58] in addition to MovieTweetings dataset (MT-750K) [59].

ML-100K dataset contains 100,000 anonymous ratings of 1682 movies made by 943 MovieLens users while ML-1M dataset consists of around 1,000,209 ratings made by 6,040 users over 3,900 movies. Both datasets contain users with more than 20 ratings given to items on a scale of 1 to 5. Lastly, the MT-750K dataset provides 742,993 ratings on a 0 to 10 scale contained in tweets posted on Twitter by 55,429 users for 32,348 movies. The basic statistics of the used datasets are summarized in Table 1.

Table 1. Statistics of the used Datasets

Dataset	ML-100K	ML-1M	MT-750K
#User	943	6,040	55,429
#Item	1,682	3,900	32,348
#Ratings	100,000	1,000,209	742,993
#Category	19	18	28
Sparsity	93,69%	95,75%	99,95%

4.2 Evaluation Procedure and Metrics

In this paper, we performed a 5-fold cross-validation over subsets generated using a sampling technique that consists of selecting randomly 80% of ratings for training and the remaining 20% of ratings to test our model. In the test set, we consider ratings as relevant items for users when $r \geq 3$ for ML-100K and ML-1M datasets and $r \geq 7$ for MT-750K dataset and the rest as not relevant.

Besides, we evaluated our proposed approach according to two topics : (1) quality and accuracy of recommendations using Precision (Prec) [60] and nDCG [61] metrics, and (2) diversity of the recommended items through Intent-aware Precision (Prec IA) [62] and α-nDCG [63] metrics.

Besides, since we seek to resolve the diversity-accuracy dilemma, we used the harmonic mean of accuracy and diversity, also known as F1-score [60]. Its is formally defined as: $F1 - score = 2 \times Accuracy \times Diversity / Accuracy + Diversity$.
In our study, all defined metrics takes values between 0 and 1, so we define two F1-score measures: The F1-nDCG measure as the trade-off between nDCG and α-nDCG and the F1-Prec measure as the trade-off between Precision and Prec_IA.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4.3 Baseline Methods

To evaluate the effectiveness of the proposed approach, we compare our method with the following baseline methods:

- **Node2Vec** (N2V): We run node2vec [41] on user-item graph constructed using evaluations made by users. The constructed graph is represented as $G_b = (U \cup I, E)$ where U is the set of users, I is the set of items and $E = \{e = r(u,i): u, i \in |U| \times |I|\}$ is a set of edges linking users and items.
- **UBCF**: The basic User-Based Collaborative filtering algorithm (UBCF) [64] finds out the users similar to the target user and then recommend items based on these similar users evaluations. We choose the Pearson correlation coefficient as the similarity measure between users, and the size of the neighborhood as 30.
- **NMF**: Non-negative matrix factorization (NMF) [65] is a dimension reduction technique adapted to sparse matrices with positive data. It attempts to approximate the input matrix with the multiplication of low-rank representations of dimensionality f.
- **TDERank**: A recommendation ranking method [66] named "Total Diversity Effect Ranking" for improving recommendation diversity. It applies the User-Based Collaborative Filtering to generate Top-(N+S) recommendations. Then, the final Top-N recommendations are proposed based on the total diversity effect of each item. As recommended by the authors, we set the S to 4 in our experiments.

In this work, we kept the default parameters on the released implementation of the authors for the baseline methods. For NMF and Node2Vec methods, we adopt the inner product of user and item embedding vectors to predict the preferences of users on unseen items.

4.4 Experimental Results

As previously mentioned, the aim of a recommender system is to generate a list of Top-N recommendations for each user. Accordingly, all measures are computed at cutoff values $N \in \{3,5,10,20\}$. In addition, the optimal value of λ parameter was identified through a grid search presented in Section 4.4.1. We set $\lambda = 0.8$.

4.4.1 Identifying the optimal value of λ parameter

The λ parameter determines the sensitivity of recommendation relevancy to the redundancy of categories. Smaller value of λ makes recommendation relevancy tolerant to category redundancies, whereas larger value have an exclusion effect of redundant categories.

In this experiment, we aim to choose the optimal value of λ parameter who leads to the best relevancy that is defined as a compromise between accuracy and diversity. For different values of λ that vary in the interval of [0,1] with a step size of 0.1, we calculate nDCG, α-nDCG, and F1-nDCG measures using our approach, averaged over the five data splits of ML-100K and ML-1M datasets at cutoff $N = 3$. The results for ML-100K and ML-1M datasets are displayed in Figure 2. For both datasets, the values of different measures increase with increasing λ up to approximately 0.8 where F1-nDCG reaches its maximum of 0.8987 at $\lambda = 0.8$ and 0.9026 at $\lambda = 0.9$ respectively for ML100K and ML-1M datasets. In general, the results show that the variation of different measures remains more or less stable after $\lambda = 0.8$. Consequently, λ will be 0.8 in our experiments, which means that a category c will be trivialized by the number of its occurrences in the pre-recommended list at k-th position according to the formula $(0.8)^{occu(c)@k}$.

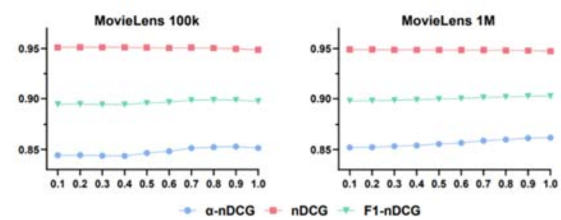


Figure 2. Effect of adjusting λ values for the ML-100K and ML-1M datasets

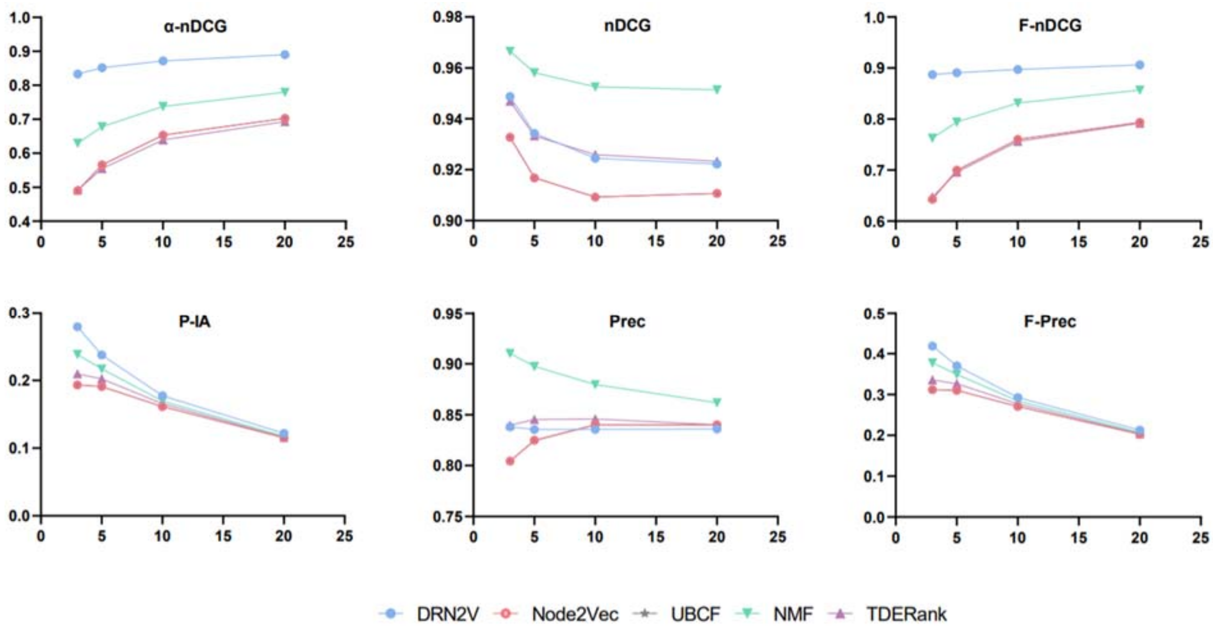
4.4.2 Results

In this section, we compare the results obtained by our approach DRN2V with those obtained by the four baseline recommendation algorithms, including the basic Node2vec, UBCF, NMF, and TDERank, as shown in Figure 3.

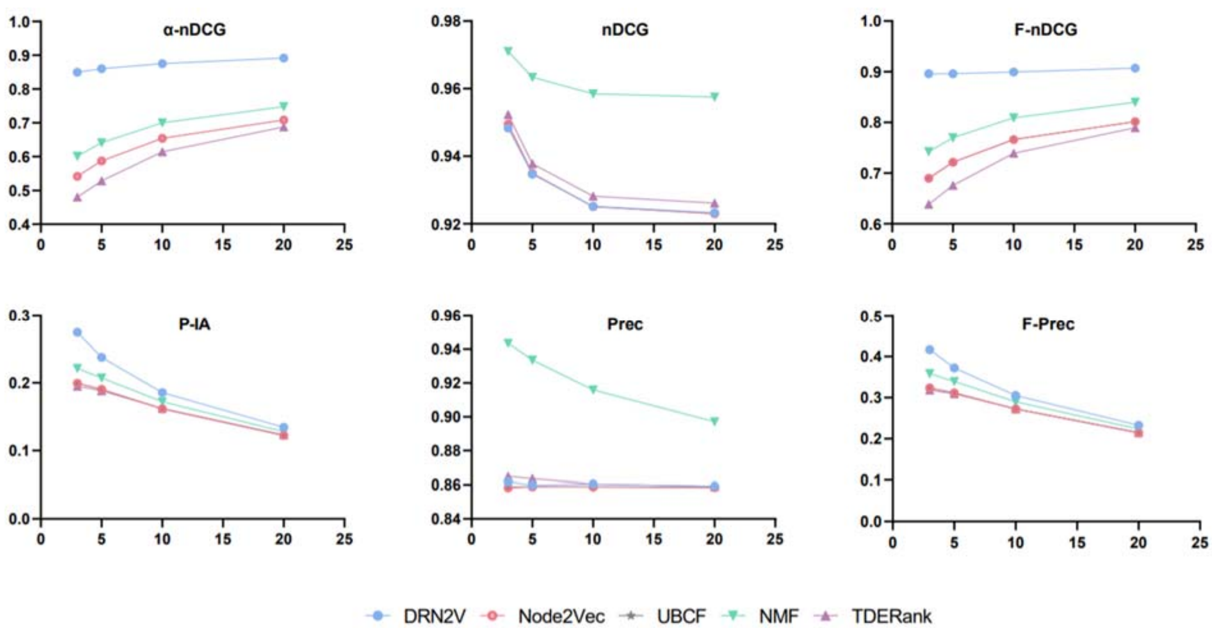
In general, in terms of diversity and F1-scores, we can see that our approach can produce more diverse recommendations with an insignificant loose in accuracy for different cutoff values. Otherwise, the NMF-based algorithm performs best in terms of accuracy since our approach doesn't consider only on providing accurate recommendations, as the NMF-based algorithm does, but it also takes into account diversity in the recommendation process which automatically generates a loss of accuracy. In spite of this disadvantage, the loss of accuracy in our approach remains minimal, whose value doesn't generally exceed 4.5% on average compared to

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

NMF-based algorithm. At the same time, the results in Figure 3 show that our approach made a significant improvement to the original Node2Vec-based algorithm whose behavior close to UBCF algorithm in most cases. For instance, the improvement compared to the original Node2Vec-based algorithm reaches 38%, 30%, and 4% on ML-100k, ML-1M, and ML-750K datasets respectively regarding the F1-nDCG@3 metrics. Besides, we can see that unlike other baseline algorithms, our algorithm remains almost stable in terms of diversity in most curves with respect to the window size (i.e. the cutoff N). As a result, our algorithm proves its capacity of adaptation to several screen sizes of devices where the system will require diversity within a sliding window.



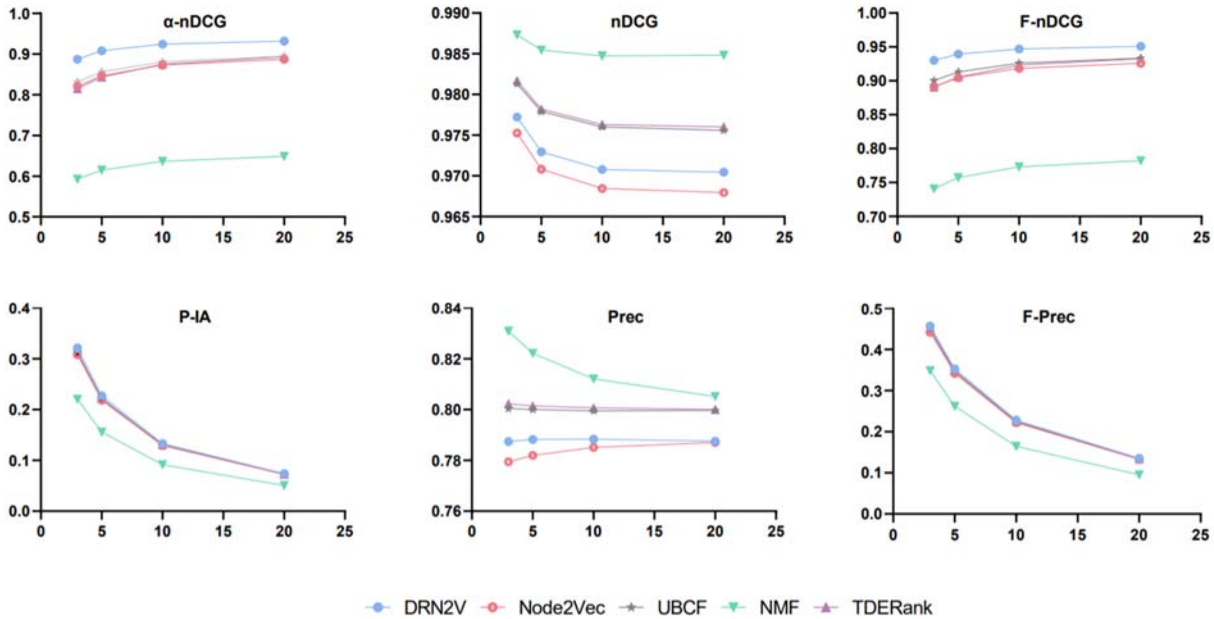
(a) MovieLens 100k



(b) MovieLens 1M

Journal of Theoretical and Applied Information Technology
15th September 2020. Vol.98. No 17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195



(c) MovieLens 750k

Figure 3. Comparison of different recommendation results of the proposed approach DRN2V and the four baselines in terms of nDCG, α-nDCG, Prec, Prec-IA, and F1-Score metrics on the three datasets: (a) ML-100K, (b) ML-1M, and (c) ML-75

To summarize the figures above, Tables 2 and 3 below show the results of different approaches by averaging different metrics obtained at cutoff of 3. The choice of the cutoff 3 comes from the fact that each item belongs to about 2 categories and the total number of categories is less than 30 (see Table 1). So, it is more likely that a list with a large number of recommended items will cover a large number of categories, which will make the diversity metrics less significant. Accordingly, we can notice the performance and robustness of our approach that show a significant improvement in diversity and F-scores.

As shown in Table 2, for the three datasets including ML-100k, ML-1M and MT-750k, our approach DRN2V and the NMF approach behaves better than the rest of other algorithms. We note that our approach has achieved greater improvement on diversity based on α-nDCG and Prec-IA metrics. While for accuracy defined by Prec and nDCG metrics, we can see that the NMF algorithm has a quite better performance. Such improvement on each metric is not at the same level for the two algorithms. As a matter of fact, if we compare the percentage of improvement between the two, we can see the huge difference that makes our approach to recommendation. For example, for the ML-100K dataset, our approach shows an improvement of 32.3% in terms of diversity (α-nDCG metric) comparing to NMF algorithm, while comparing to our approach DRN2V, the NMF algorithm showed an improvement of 1.8% in terms of accuracy (nDCG metric).

Additionally, the main focus of our approach is to maintain a good level of both accuracy and diversity. In accordance, we mainly focus, in the following, on the performance of different algorithms regarding the combination of accuracy and diversity using the F1-score metric as shown in Table 3. As we can see in Table 3, for the three datasets, i.e., ML-100K, ML-1M and ML-750K, our approach DRN2V presents the best compromise between accuracy and diversity in all cases, by adopting any baseline recommendation algorithm and using any performance F1-score measurement.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 2. The average performances of the proposed approach DRN2V and the four baselines on the three datasets at cutoff N=3 regarding Prec, nDCG, α-nDCG and Prec IA metrics

Model	Prec	nDCG	α-nDCG	Prec- IA
		ML-100K		
DRN2V	0,8381	0,9487	**0,8332**	**0,2793**
N2V	0,8044	0,9327	0,4904	0,1933
UBCF	0,7985	0,9356	0,4855	0,1913
NMF	**0,9102**	**0,9665**	0,6298	0,2382
TDRank	0,8396	0,9469	0,4917	0,2097
		ML-1M		
DRN2V	0,8617	0,9483	**0,8496**	**0,2748**
N2V	0,8582	0,9494	0,5416	0,1992
UBCF	0,8584	0,9494	0,5417	0,1993
NMF	**0,9434**	**0,9710**	0,6011	0,2214
TDERank	0,8648	0,9523	0,4802	0,1952
		MT-750K		
DRN2V	0,7874	0,9772	**0,8874**	**0,3216**
N2V	0,7794	0,9753	0,8210	0,3082
UBCF	0,8000	0,9811	0,8313	0,3118
NMF	**0,8309**	**0,9873**	0,5925	0,2205
TDERank	0,8023	0,9817	0,8151	0,3135

Table 3. The average performances of the proposed approach DRN2V and the four baselines on the three datasets at cutoff N=3 regarding F1-score metric

F-score	F1-nDCG	F1-Prec
	ML-100K	
DRN2V	**0,8872**	**0,4189**
N2V	0,6428	0,3116
UBCF	0,6392	0,3086
NMF	0,7626	0,3775
TDERank	0,6473	0,3356
	ML-1M	
DRN2V	**0,8962**	**0,4167**
N2V	0,6897	0,3233
UBCF	0,6898	0,3234
NMF	0,7425	0,3586
TDERank	0,6385	0,3185
	MT-750K	
DRN2V	**0,9302**	**0,4567**
N2V	0,8915	0,4417
UBCF	0,9000	0,4487
NMF	0,7406	0,3485
TDERank	0,8907	0,4508

At last, although generally, recommendation algorithms tend to improve diversity causes a loss on accuracy, our method has shown its efficiency in improving diversity while maintaining a good level of accuracy even in high sparse data. Consequently, our approach has shown its potential to be widely applied in large-scale recommender systems.

5. CONCLUSION

In this paper, we propose DRN2V, a new approach based on NE algorithm that aims to solve the accuracy-diversity dilemma. Primarily, we build the User-Item and the Item-Category subgraphs to use them as input for the node2vec technology to extract the representations of each node while respecting the structure of the network. Besides, we propose a new formula to predict the items evaluations based on user preferences and the redundancy of categories in recommended list. Finally, we compare our approach with several embedding and user-based recommendation algorithms on three real datasets. The results show that our approach outperforms the baselines algorithms while respecting the compromise between accuracy and diversity. Future work will involve a more comprehensive evaluation, trying to find the optimal λ parameter using some optimization algorithms, as well as evaluating our model performance in a real recommendation scenario. Furthermore, we are interested in utilizing other Network Embedding's methods using more complex networks, other than the node2vec algorithm, to improve recommendation diversity.

REFRENCES:

[1] W. Hill, L. Stead, M. Rosenstein, and G. Furnas, "Recommending and evaluating choices in a virtual community of use," in Proceedings of the SIGCHI conference on Human factors in computing systems, 1995, pp. 194–201.

[2] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "GroupLens: an open architecture for collaborative filtering of netnews," in Proceedings of the 1994 ACM conference on Computer supported cooperative work, 1994, pp. 175–186.

[3] U. Shardanand and P. Maes, "Social information filtering: algorithms for automating" word of mouth"," in Chi, 1995, Vol. 95, pp. 210–217.

[4] A. Gholami, Y. Forghani, and M. Branch, "Improving Multi-class Co-Clustering-Based

Collaborative Recommendation Using Item Tags," Revue d'Intelligence Artificielle, Vol. 34, No. 1, pp. 59–65, 2020, doi: https://doi.org/10.18280/ria.340108.

[5] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems," ACM Transactions on Information Systems (TOIS), Vol. 22, No. 1, pp. 5–53, 2004.

[6] B. Smyth and P. McClave, "Similarity vs. diversity," in International conference on case-based reasoning, 2001, pp. 347–361.

[7] M. Ge, C. Delgado-Battenfeld, and D. Jannach, "Beyond accuracy: evaluating recommender systems by coverage and serendipity," in Proceedings of the fourth ACM conference on Recommender systems, 2010, pp. 257–260.

[8] A. Landin, E. Suárez-García, and D. Valcarce, "When Diversity Met Accuracy: A Story of Recommender Systems," Multidisciplinary Digital Publishing Institute Proceedings, Vol. 2, No. 18, p. 1178, 2018.

[9] S. M. McNee, J. Riedl, and J. A. Konstan, "Being accurate is not enough: how accuracy metrics have hurt recommender systems," in CHI'06 extended abstracts on Human factors in computing systems, 2006, pp. 1097–1101.

[10] K. Bradley and B. Smyth, "Improving recommendation diversity," in Proceedings of the Twelfth Irish Conference on Artificial Intelligence and Cognitive Science, Maynooth, Ireland, 2001, pp. 85–94.

[11] C.-N. Ziegler, S. M. McNee, J. A. Konstan, and G. Lausen, "Improving recommendation lists through topic diversification," in Proceedings of the 14th international conference on World Wide Web, 2005, pp. 22–32.

[12] S. Vargas and P. Castells, "Rank and Relevance in Novelty and Diversity Metrics for Recommender Systems," in Proceedings of the Fifth ACM Conference on Recommender Systems, New York, NY, USA, 2011, pp. 109–116, doi: 10.1145/2043932.2043955.

[13] Y. C. Zhang, D. Ó. Seaghdha, D. Quercia, and T. Jambor, "Auralist: introducing serendipity into music recommendation," in Proceedings of the fifth ACM international conference on Web search and data mining, 2012, pp. 13–22.

[14] P. Cheng, S. Wang, J. Ma, J. Sun, and H. Xiong, "Learning to Recommend Accurate and Diverse Items," in Proceedings of the 26th International Conference on World Wide Web - WWW '17, Perth, Australia, 2017, pp. 183–192, doi: 10.1145/3038912.3052585.

[15] N. Hurley and M. Zhang, "Novelty and Diversity in Top-N Recommendation – Analysis and Evaluation," ACM Trans. Internet Technol., Vol. 10, No. 4, p. 14:1–14:30, Mar. 2011, doi: 10.1145/1944339.1944341.

[16] S. Vargas, L. Baltrunas, A. Karatzoglou, and P. Castells, "Coverage, redundancy and size-awareness in genre diversity for recommender systems," in Proceedings of the 8th ACM Conference on Recommender systems - RecSys '14, Foster City, Silicon Valley, California, USA, 2014, pp. 209–216, doi: 10.1145/2645710.2645743.

[17] Q. Wang, J. Yu, and W. Deng, "An adjustable re-ranking approach for improving the individual and aggregate diversities of product recommendations," Electronic Commerce Research, Vol. 19, No. 1, pp. 59–79, Mar. 2019, doi: 10.1007/s10660-018-09325-4.

[18] G. Adomavicius and Y. Kwon, "Overcoming accuracy-diversity tradeoff in recommender systems: a variance-based approach," in Proceedings of WITS, Vol. 8, 2008.

[19] A. Gogna and A. Majumdar, "DiABlO: Optimization based design for improving diversity in recommender system," Information Sciences, Vol. 378, pp. 59–74, 2017.

[20] E. M. Hamedani and M. Kaedi, "Recommending the long tail items through personalized diversification," Knowledge-Based Systems, Vol. 164, 2019, pp. 348 – 357, doi: https://doi.org/10.1016/j.knosys.2018.11.004.

[21] A. Pathak and B. K. Patra, "A knowledge reuse framework for improving novelty and diversity in recommendations," in Proceedings of the Second ACM IKDD Conference on Data Sciences, 2015, pp. 11–19.

[22] L. Yezheng, W. Jinkun, J. Yuanchun, S. Jianshan, and S. Chunhua, "Utilize Item Correlation to Improve Aggregate Diversity for Recommender Systems," in 2016 IEEE First International Conference on Data Science in Cyberspace (DSC), 2016, pp. 412–417.

[23] G. Adomavicius and Y. Kwon, "Maximizing Aggregate Recommendation Diversity: A Graph-Theoretic Approach," p. 8.

[24] G. Adomavicius and YoungOk Kwon, "Improving Aggregate Recommendation

Diversity Using Ranking-Based Techniques," IEEE Transactions on Knowledge and Data Engineering, Vol. 24, No. 5, May 2012, pp. 896–911, doi: 10.1109/TKDE.2011.15.

[25] J. Díez, D. Martínez-Rego, A. Alonso-Betanzos, O. Luaces, and A. Bahamonde, "Optimizing novelty and diversity in recommendations," Progress in Artificial Intelligence, Jul. 2018, doi: 10.1007/s13748-018-0158-4.

[26] J. Wang, Y. Liu, J. Sun, Y. Jiang, and C. Sun, "Diversified Recommendation Incorporating Item Content Information Based on MOEA/D," in 2016 49th Hawaii International Conference on System Sciences (HICSS), Koloa, HI, USA, Jan. 2016, pp. 688–696, doi: 10.1109/HICSS.2016.91.

[27] A. Ashkan, B. Kveton, S. Berkovsky, and Z. Wen, "Optimal Greedy Diversity for Recommendation," 2015, p. 7.

[28] J. Borràs, A. Moreno, and A. Valls, "Diversification of recommendations through semantic clustering," Multimedia Tools and Applications, Vol. 76, No. 22, Nov. 2017, pp. 24165–24201, doi: 10.1007/s11042-016-4166-8.

[29] T. Di Noia, J. Rosati, P. Tomeo, and E. D. Sciascio, "Adaptive multi-attribute diversity for recommender systems," Information Sciences, Vol. 382–383, Mar. 2017, pp. 234–253, doi: 10.1016/j.ins.2016.11.015.

[30] M. Kaminskas and D. Bridge, "Diversity, Serendipity, Novelty, and Coverage: A Survey and Empirical Analysis of Beyond-Accuracy Objectives in Recommender Systems," ACM Transactions on Interactive Intelligent Systems, Vol. 7, No. 1, Dec. 2016, pp. 1–42, doi: 10.1145/2926720.

[31] F. Eskandanian, B. Mobasher, and R. Burke, "A Clustering Approach for Personalizing Diversity in Collaborative Recommender Systems," in Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization - UMAP '17, Bratislava, Slovakia, 2017, pp. 280–284, doi: 10.1145/3079628.3079699.

[32] S. Li, Y. Zhou, D. Zhang, Y. Zhang, and X. Lan, "Learning to Diversify Recommendations Based on Matrix Factorization," in 2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech), Orlando, FL, Nov. 2017, pp. 68–74, doi: 10.1109/DASC-PICom-DataCom-CyberSciTec.2017.26.

[33] S. Vargas and P. Castells, "Exploiting the Diversity of User Preferences for Recommendation," in Proceedings of the 10th conference on open research areas in information retrieval, 2013, pp. 129–136.

[34] X. Xing, C. Sha, and J. Niu, "Improving Topic Diversity in Recommendation Lists: Marginally or Proportionally?" in Web and Big Data, Vol. 10367, L. Chen, C. S. Jensen, C. Shahabi, X. Yang, and X. Lian, Eds. Cham: Springer International Publishing, 2017, pp. 142–150.

[35] M. Zanitti, S. Kosta, and J. Sørensen, "A User-Centric Diversity by Design Recommender System for the Movie Application Domain," in Companion of the The Web Conference 2018 on The Web Conference 2018 - WWW '18, Lyon, France, 2018, pp. 1381–1389, doi: 10.1145/3184558.3191580.

[36] S. Zhang, Y. Xia, and X. Li, "Diversified recommendation algorithm for hybrid label based on matrix factorization," in 2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), Chengdu, Apr. 2018, pp. 39–44, doi: 10.1109/ICCCBDA.2018.8386484.

[37] C. C. Aggarwal, Recommender systems: the textbook, 1st edition. New York, NY: Springer Science+Business Media, 2016.

[38] D. Zhang, J. Yin, X. Zhu, and C. Zhang, "Network Representation Learning: A Survey," arXiv:1801.05852 [cs, stat], Jul. 2018, Accessed: Dec. 17, 2019. [Online]. Available: http://arxiv.org/abs/1801.05852.

[39] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed Representations of Words and Phrases and their Compositionality," p. 9.

[40] B. Perozzi, R. Al-Rfou, and S. Skiena, "DeepWalk: online learning of social representations," in Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '14, New York, New York, USA, 2014, pp. 701–710, doi: 10.1145/2623330.2623732.

[41] A. Grover and J. Leskovec, "node2vec: Scalable Feature Learning for Networks," arXiv:1607.00653 [cs, stat], Jul. 2016, Accessed: Dec. 18, 2019. [Online]. Available: http://arxiv.org/abs/1607.00653.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[42] Y. Bengio, A. Courville, and P. Vincent, "Representation Learning: A Review and New Perspectives," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 35, No. 8, Aug. 2013, pp. 1798–1828, doi: 10.1109/TPAMI.2013.50.

[43] E. Palumbo, G. Rizzo, and R. Troncy, "entity2rec: Learning User-Item Relatedness from Knowledge Graphs for Top-N Item Recommendation," in Proceedings of the Eleventh ACM Conference on Recommender Systems - RecSys '17, Como, Italy, 2017, pp. 32–36, doi: 10.1145/3109859.3109889.

[44] E. Palumbo, G. Rizzo, R. Troncy, E. Baralis, M. Osella, and E. Ferro, "Knowledge Graph Embeddings with node2vec for Item Recommendation," in The Semantic Web: ESWC 2018 Satellite Events, Vol. 11155, A. Gangemi, A. L. Gentile, A. G. Nuzzolese, S. Rudolph, M. Maleshkova, H. Paulheim, J. Z. Pan, and M. Alam, Eds. Cham: Springer International Publishing, 2018, pp. 117–120.

[45] L. Xiaowen, W. Zijian, and D. Bosen, "vec2rec: Network Embedding for Item-to-Item Recommendation," Stanford CS224W Project, 2018. [Online]. Available: http://snap.stanford.edu/class/cs224w-2018/reports/CS224W-2018-14.pdf.

[46] L. Liang and R. Tang, "An Improved Collaborative Filtering Algorithm Based on Node2vec," in Proceedings of the 2018 2nd International Conference on Computer Science and Artificial Intelligence - CSAI '18, Shenzhen, China, 2018, pp. 218–222, doi: 10.1145/3297156.3297219.

[47] J. Chen et al., "N2VSCDNNR: A Local Recommender System Based on Node2vec and Rich Information Network," arXiv:1904.12605 [physics], Apr. 2019, Accessed: Sep. 05, 2019. [Online]. Available: http://arxiv.org/abs/1904.12605.

[48] T. Deselaers, T. Gass, P. Dreuw, and H. Ney, "Jointly optimising relevance and diversity in image retrieval," in Proceedings of the ACM international conference on image and video retrieval, 2009, p. 39.

[49] M. Drosou and E. Pitoura, "Diversity over Continuous Data.," IEEE Data Eng. Bull., Vol. 32, No. 4, 2009, pp. 49–56.

[50] R. Boim, T. Milo, and S. Novgorodov, "Diversification and refinement in collaborative filtering recommender," in Proceedings of the 20th ACM international conference on Information and knowledge management, 2011, pp. 739–744.

[51] X. Li and T. Murata, "Using multidimensional clustering based collaborative filtering approach improving recommendation diversity," in 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, Vol. 3, 2012, pp. 169–174.

[52] L. Shi, "Trading-off among accuracy, similarity, diversity, and long-tail: a graph-based recommendation approach," in Proceedings of the 7th ACM conference on Recommender systems, 2013, pp. 57–64.

[53] M. Zhang and N. Hurley, "Novel Item Recommendation by User Profile Partitioning," in Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology - Vol. 01, Washington, DC, USA, 2009, pp. 508–515, doi: 10.1109/WI-IAT.2009.85.

[54] B. Cai, X. Yang, Y. Huang, H. Li, and Q. Sang, "A Triangular Personalized Recommendation Algorithm for Improving Diversity," Discrete Dynamics in Nature and Society, Vol. 2018, pp. 1–11, 2018, doi: 10.1155/2018/3162068.

[55] J. Wasilewski and N. Hurley, "Intent-aware diversification using a constrained PLSA," in Proceedings of the 10th ACM Conference on Recommender Systems, 2016, pp. 39–42.

[56] V. Pouli, J. S. Baras, and A. Arvanitis, "Increasing recommendation accuracy and diversity via social networks hyperbolic embedding," in 2014 IEEE 11th Consumer Communications and Networking Conference (CCNC), Las Vegas, NV, Jan. 2014, pp. 225–232, doi: 10.1109/CCNC.2014.6866575.

[57] J. Lin et al., "RNE: A Scalable Network Embedding for Billion-Scale Recommendation," in Advances in Knowledge Discovery and Data Mining, Vol. 11440, Q. Yang, Z.-H. Zhou, Z. Gong, M.-L. Zhang, and S.-J. Huang, Eds. Cham: Springer International Publishing, 2019, pp. 432–445.

[58] F. M. Harper and J. A. Konstan, "The movielens datasets: History and context," Acm transactions on interactive intelligent systems (tiis), Vol. 5, No. 4, p. 19, 2016.

[59] S. Dooms, T. De Pessemier, and L. Martens, "Movietweetings: a movie rating dataset collected from twitter," in Workshop on Crowdsourcing and human computation for recommender systems, CrowdRec at RecSys, 2013, Vol. 2013, p. 43.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[60] D. M. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," 2011.

[61] K. Järvelin and J. Kekäläinen, "Cumulated gain-based indicators of IR performance," p. 26, 2002.

[62] R. Agrawal, S. Gollapudi, A. Halverson, and S. Ieong, "Diversifying Search Results," in Proceedings of the Second ACM International Conference on Web Search and Data Mining, New York, NY, USA, 2009, pp. 5–14, doi: 10.1145/1498759.1498766.

[63] C. L. A. Clarke et al., "Novelty and diversity in information retrieval evaluation," in Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '08, Singapore, Singapore, 2008, p. 659, doi: 10.1145/1390334.1390446.

[64] J. Herlocker, "An Empirical Analysis of Design Choices in Neighborhood-Based Collaborative Filtering Algorithms," p. 24.

[65] C. Ding, T. Li, and M. Jordan, "Convex and Semi-Nonnegative Matrix Factorizations," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 1, No. 32, 2010, pp. 45–55.

[66] W. Premchaiswadi, P. Poompuang, N. Jongswat, and N. Premchaiswadi, "Enhancing Diversity-Accuracy Technique on User-Based Top-N Recommendation Algorithms," in 2013 IEEE 37th Annual Computer Software and Applications Conference Workshops, Japan, Jul. 2013, pp. 403–408, doi: 10.1109/COMPSACW.2013.68.