Latent Modeling of Unexpectedness for Recommendations

Pan Li
New York University
New York, NY, USA
pli2@stern.nyu.edu

Alexander Tuzhilin
New York University
New York, NY, USA
atuzhili@stern.nyu.edu

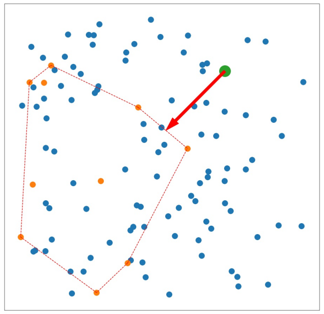

Figure 1: Visualization of Latent Convex Hull. Unexpectedness is defined as the distance between new item and latent convex hull generated by all consumed items.

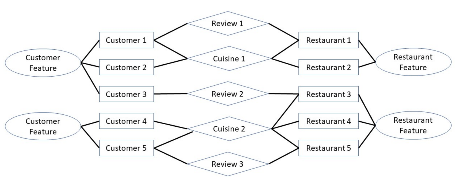

Figure 2: Visualization of Heterogeneous Information Network

ABSTRACT

Unexpectedness constitutes an important factor for recommender system to improve user satisfaction and avoid filter bubble issues. Previous methods model unexpectedness in the *feature space*, making them difficult to capture the latent, complex and heterogeneous interactions between users and items. In this paper, we propose to model unexpectedness in the *latent space* and utilize a latent convex hull structure to provide unexpected recommendations, as illustrated in Figure 1. Extensive experiments on two real-world datasets demonstrate effectiveness of latent unexpectedness over explicit unexpectedness and show that the proposed model significantly outperforms baseline models in terms of unexpectedness measures while achieving the same level of accuracy.

KEYWORDS

Unexpected Recommendation, Latent Embeddings, Heterogeneous Information Network

INTRODUCTION AND RELATED WORK

Recommender systems have been playing an important role in assisting users in filtering for the best content and shaping their consumption behaviors. To solve the problem of filter bubbles and boredom [11, 12], researchers introduce the measure of unexpectedness [10, 14] beyond accuracy, the goal of which is to provide novel and not previously seen recommendations. It is positively correlated with user satisfaction and helps to achieve strong recommendation performance [3–5, 8].

However, one limitation with unexpectedness is that it is defined in the *feature space* of previously consumed items, which relies completely on explicit purchase information, and may not work well

ACM RecSys 2019 Late-breaking Results, 16th-20th September 2019, Copenhagen, Denmark

Evaluation Metrics

RMSE & MAE
Root Mean Square Error and Mean Absolute Error measures the differences between estimated ratings and actual ratings.

Precision@N & Recall@N
Precision is the fraction of the recommended items that are of high utility to the user. Recall is the fraction of the high utility items that are recommended to the user.

Unexpectedness
Unexpectedness is calculated through equation (3) following our proposed definition.

Serendipity
Serendipity [8] is computed by $Serendipity = \frac{RS\&PM\&Useful}{RS}$ where RS stands for the recommended items using the target model, PM stands for the recommendation items using a primitive prediction algorithm and USEFUL stands for the items whose utility is above certain threshold.

Diversity
Diversity is computed as the average intra-list distance [17]: $D(R) = \sum_{i \in R} \sum_{j \neq i \in R} d(i, j)$

Coverage
Coverage is computed as the percentage of distinctive recommended items over all distinctive items in the dataset.

in the case when purchase actions are sparse or noisy. Besides, it does not include specific user and item features, neither does it address the latent and complex relationships between users and items. Furthermore, the distance metric between discrete items is hard to define in the feature space, while in the latent space we have well-defined euclidean distances. Thus, we cannot model the unexpected relations efficiently in the feature space.

Therefore, though researchers have successfully developed algorithms to improve unexpectedness and other novelty metrics, they always come at a price of losing accuracy performance, as pointed out in [19]. This severely limits the practical use of the unexpected recommendation, especially in business applications where the goal is to increase commercial sales and enhance user satisfaction. Thus, it is important to design a novel unexpected recommender to increase novelty of recommendations while keeping the same level of accuracy performance.

In this paper, we propose to define unexpectedness as the distance metric in the *latent space* (latent feature and attribute embeddings) rather than *feature space* (explicit users and items) by utilizing Heterogeneous Information Network Embeddings (HINE) [15, 16] to capture the latent, complex and heterogeneous relations between them. We take the natural closure of convex hull in the latent space, and calculate unexpectedness as the euclidean distance from new item embedding to the latent convex hull of expected items of the user. The proposed unexpectedness measure is subsequently combined with estimated ratings for providing recommendations.

We make the following contributions in this paper. We propose to model unexpectedness in the latent space by defining the expected set for each user as a latent convex hull generated by item embeddings. We also conduct extensive experiments on two real-world datasets and show that our model significantly improves novelty measures without losing any accuracy metrics.

METHOD

In prior literature [4], unexpectedness is defined as the distance of recommended item from the closure set of expected items, the latter including items that either previously consumed by the user or closely related to the consumptions. However, this definition does not include the feature information for users and items, neither does it consider the latent and heterogeneous interactions between them. Thus, it might not serve as a perfect definition, for we cannot manually codify all explicit associations and relations between users and items. Also, it is hard to mathematically formalize the expected set in the feature space, where the distribution of users and items is discrete and unstructured.

On the other hand, heterogeneous information network [16] (HIN) along with latent embedding method [7] provides us with an efficient tool to model users, items and their associated features simultaneously by linking the associated features with corresponding entities in the network and subsequently map them into the latent space. Thus in the paper, we propose to combine those two techniques for latent modeling of unexpectedness.

Baseline Models

SPR [9]
Serendipitous Personalized Ranking extends traditional personalized ranking methods by considering item popularity in AUC optimization, which makes the ranking sensitive to the popularity of negative examples.

Auralist [18]
Auralist is a personalized recommendation system that balances between the desired goals of accuracy and novelty simultaneously.

HOM-LIN [4]
HOM-LIN is the state-of-the-art unexpected algorithm that utilizes the combination of estimated ratings and unexpectedness for recommendation.

DPP [6]
The Determinantal Point Process utilizes a fast greedy MAP inference approach to generate relevant and diverse recommendations.

Random
Random is the baseline model where we randomly recommend items to users without considering any information about the ratings, unexpectedness and utility.

We denote the heterogeneous network as $G = (V, E, T)$, in which each node v and each link e are assigned with specific type T_v and T_e. To effectively learn node representations, we enable the skip-gram mechanism to maximize the probability of each context node c_t within the neighbors of v, denoted as $N_t(v)$, where we add the subscript t ($t \in T_v$) to limit the node to a specific type:

$$arg\,max_\theta \sum_{v \in V} \sum_{t \in T_v} \sum_{c_t \in N_t(v)} logP(c_t|v;\theta) \tag{1}$$

To calculate $P(c_t|v;\theta)$, we propose to use heterogeneous random walk to generate meta-paths of multiple types of nodes in the form of $V_1 \xrightarrow{R_1} V_2 \xrightarrow{R_2} V_3 \cdots V_n$ wherein $R = R_1 \circ R_2 \circ \cdots R_n$ defines the composite relations between the start and the end of the heterogeneous random walk. The transition probability within each random walk between two nodes is defined as follows:

$$p(V_{t+1}|V_t) = \begin{cases} \frac{C(T_{V_t}, T_{V_{t+1}})}{|N_{t+1}(V_t)|}, & (V_t, V_{t+1}) \in E \\ 0, & (V_t, V_{t+1}) \notin E \end{cases} \tag{2}$$

where $C(T_{V_t}, T_{V_{t+1}})$ stands for the transition coefficient between the type of node V_t and the type of node V_{t+1}. $|N_{t+1}(V_t)|$ stands for the number of nodes of type V_{t+1} in the neighborhood of V_t. We apply heterogeneous random walk iteratively to each node and generate the collection of meta-path sequences for the skip-gram mechanism.

Note that in previous definition [4], the idea of ‘’closure set” plays an important role. As a natural closure in the latent space, latent convex hull of each user is the smallest convex set that contains all the item embeddings that the user has visited. Under this setting, we assume that the expected set maintains its convexity in the growing process. For example, if a person enjoy rap music and rock & roll, she/he shall also have certain expectations on rap-rock music, an innovation combination of those two genres. Subsequently, we define the unexpectedness as **the distance between the embedding of new item and the latent convex hull generated from the embeddings of all visited items of the user**:

$$Unexp_{u,i} = d(i; LC(N_i)) \tag{3}$$

where $N_i = (i_1, i_2, \cdots, i_n)$ contains the embeddings of items that link to the user in the heterogeneous information network. Specifically, the distance function is well defined as the minimal distance from the given point to all the boundaries of the latent closure. We visualize this definition in Figure 1.

Once we set up the definition of unexpectedness, we perform the unexpected recommendation using hybrid-utility based collaborative filtering methods that linearly combine estimated ratings (representing accuracy) with unexpectedness (representing novelty)

$$Utility_{u,i} = (1-\alpha) * EstRating_{u,i} + \alpha * Unexp_{u,i} \tag{4}$$

Dataset	Yelp	TA
# Review	5,996,996	878,561
# Business	188,593	576,689
# User	1,518,169	3,945
# Sparsity	0.002%	0.039%

Table 1: Descriptive Statistics for the two Datasets.

Data	Model	Unexp	Ser	Div	Cov
Yelp	FM	0.0326	0.0978	0.0135	0.5369
	FM+Unexp	**0.1122***	**0.4793***	**0.3798***	0.5904
	CoCluster	0.0338	0.4595	0.3106	0.5811
	CoCluster+Unexp	**0.1400***	**0.4660***	**0.3269***	0.5904
	SVD	0.0457	0.1352	0.0479	0.5221
	SVD+Unexp	**0.0620***	**0.1469***	**0.0524***	0.5406
	NMF	0.0333	0.4954	0.3268	0.5867
	NMF+Unexp	**0.1390***	**0.5469***	**0.3430***	0.5904
	KNN	0.0448	0.0977	0.0129	0.5369
	KNN+Unexp	**0.0610***	**0.1165***	**0.0259***	0.5406
	SPR	0.0668	0.3720	0.2532	0.5697
	Auralist	0.0663	0.3637	0.2047	0.5457
	HOM-LIN	0.0751	0.4329	0.3011	0.5365
	DPP	0.0670	0.4488	0.2488	0.5904
	Random	0.1733	0.4848	0.3763	0.5457
TA	FM	0.0222	0.3979	0.0017	0.1798
	FM+Unexp	**0.0643***	**0.4631***	**0.0493***	0.1798
	CoCluster	0.0234	0.3973	0.0015	0.1855
	CoCluster+Unexp	**0.0652***	**0.4619***	**0.0471***	0.1807
	SVD	0.0231	0.3967	0.0006	0.1798
	SVD+Unexp	**0.0644***	**0.4621***	**0.0499***	0.1798
	NMF	0.0227	0.3979	0.0010	0.1798
	NMF+Unexp	**0.0644***	**0.4627***	**0.0499***	0.1798
	KNN	0.0233	0.3979	0.0019	0.1798
	KNN+Unexp	**0.0643***	**0.4631***	**0.0492***	0.1798
	SPR	0.0474	0.3593	0.0375	0.1834
	Auralist	0.0473	0.3462	0.0355	0.1834
	HOM-LIN	0.0572	0.3729	0.0411	0.1807
	DPP	0.0464	0.3245	0.0311	0.1807
	Random	0.0833	0.4468	0.0650	0.1834

Table 2: Comparison of recommendation performance in novelty measures, "*" stands for 95% statistical significance

The key idea lies in that, instead of recommending the similar items that the users are very familiar with as the classical recommenders do, we recommend unexpected and useful items to the users that they might have not thought about, but indeed fit well to their satisfactions. These two adversarial forces of accuracy and novelty work together to obtain the optimal solution, thus improving recommendation performance and user satisfaction. In real practice, the unexpected coefficient α is optimized through Bayesian optimization.

EXPERIMENTS AND RESULTS

We implement our model on two real-world datasets: the Yelp Challenge Dataset Round 12 [2] and the TripAdvisor Dataset [1]. The descriptive statistics of these two datasets are listed in Table 1. To address the cold-start issue, we filter out users and items that appear less than 5 times in the dataset. We measure the recommendation results using accuracy and novelty metrics listed in page 2.

We compare the recommendation performance with and without considering the proposed unexpectedness by 5-fold cross-validation experiments using five popular collaborative filtering algorithms including KNN, SVD, CoClustering, NMF and FM [13]. We also report results for baseline unexpected recommendation models. As shown in Table 2 and 3, when considering unexpectedness in the recommendation process, we obtain significant amount of improvements in terms of Unexpectedness, Serendipity and Diversity measures, without sacrificing any accuracy performance in RMSE, MAE, Precision and Recall. Therefore, the proposed definition of unexpectedness enables us to provide unexpected and useful recommendations at the same time. It is crucial for successful deployments of unexpected recommendation models in industrial applications. In addition, the proposed approach outperforms all previous unexpected recommendation models introduced in page 3, which validates the superiority of modeling unexpectedness in the latent space over feature space. Finally, we point out that the proposed model achieves greater improvements on Yelp dataset, which contains richer feature information compared to the TripAdvisor dataset. This observation is in line with our assumption that feature information matters in the definition of unexpectedness.

CONCLUSION

In this paper, we propose to model unexpectedness and user expectations in the latent space, which makes it possible to capture the latent, complex and heterogeneous relations between users and items, thus significantly improving the performance and practicability of unexpected recommendations. We empirically demonstrate that the proposed approach consistently and significantly outperforms baseline models in terms of unexpected measures without sacrificing the performance of accuracy.

As the future work, we plan to conduct live experiments with real business environment in order to further evaluate the effectiveness of unexpected recommendations and analyze both qualitative and quantitative aspects in an online retail setting, especially with the utilization of A/B test.

Data	Model	RMSE	MAE	Pre@5	Rec@5
Yelp	FM	0.9197	0.6815	0.7699	0.6123
	FM+Unexp	0.9178	0.6820	0.7700	0.6123
	CoCluster	0.9499	0.7140	0.7255	0.5913
	CoCluster+Unexp	0.9504	0.7138	0.7196	0.5864
	SVD	0.9132	0.7071	0.7692	0.5944
	SVD+Unexp	0.9134	0.7076	0.7701	0.5975
	NMF	0.9533	0.7181	0.7197	0.5318
	NMF+Unexp	0.9522	0.7154	0.7222	0.5833
	KNN	0.9123	0.7048	0.7687	0.6085
	KNN+Unexp	0.9128	0.7051	0.7659	0.6073
	SPR	1.0351	0.7729	0.7692	0.6188
	Auralist	1.0377	0.7799	0.7678	0.6000
	HOM-LIN	0.9609	0.7447	0.7621	0.6150
	DPP	1.0288	0.7702	0.7598	0.6012
	Random	1.4959	1.2456	0.4250	0.3333
TA	FM	1.1105	0.8340	0.6768	0.9590
	FM+Unexp	1.1275	0.8445	0.7040	0.9656
	CoCluster	1.0178	0.7643	0.6845	0.9732
	CoCluster+Unexp	1.0285	0.7541	0.6865	0.9703
	SVD	0.9868	0.7533	0.7210	0.9465
	SVD+Unexp	0.9937	0.7517	0.7085	0.9594
	NMF	1.0241	0.7709	0.6850	0.9681
	NMF+Unexp	1.0262	0.7533	0.6881	0.9775
	KNN	0.9940	0.7531	0.6969	0.9689
	KNN+Unexp	1.0001	0.7483	0.6907	0.9763
	SPR	1.0328	0.8008	0.6395	0.9325
	Auralist	1.0318	0.7997	0.6460	0.9390
	HOM-LIN	1.0298	0.7902	0.6420	0.9418
	DPP	1.0304	0.8158	0.6264	0.9303
	Random	1.6857	1.3100	0.3238	0.2500

Table 3: Comparison of recommendation performance in accuracy measures, "*" stands for 95% statistical significance

REFERENCES

[1] 2019. TripAdvisor Dataset. http://www.cs.cmu.edu/~jiweil/html/hotel-review.html.

[2] 2019. Yelp Challenge Dataset. https://www.yelp.com/dataset/challenge.

[3] Panagiotis Adamopoulos. 2014. On discovering non-obvious recommendations: Using unexpectedness and neighborhood selection methods in collaborative filtering systems. In *Proceedings of the 7th WSDM*.

[4] Panagiotis Adamopoulos and Alexander Tuzhilin. 2015. On unexpectedness in recommender systems: Or how to better expect the unexpected. *ACM Transactions on Intelligent Systems and Technology (TIST)* (2015).

[5] Li Chen, Yonghua Yang, Ningxia Wang, Keping Yang, and Quan Yuan. 2019. How Serendipity Improves User Satisfaction with Recommendations? A Large-Scale User Evaluation. In *The World Wide Web Conference*. ACM, 240–250.

[6] Laming Chen, Guoxin Zhang, and Eric Zhou. 2018. Fast greedy MAP inference for Determinantal Point Process to improve recommendation diversity. In *Advances in Neural Information Processing Systems*. 5622–5633.

[7] Yuxiao Dong, Nitesh V Chawla, and Ananthram Swami. 2017. metapath2vec: Scalable representation learning for heterogeneous networks. In *Proceedings of the 23rd ACM SIGKDD*. ACM.

[8] Mouzhi Ge, Carla Delgado-Battenfeld, and Dietmar Jannach. 2010. Beyond accuracy: evaluating recommender systems by coverage and serendipity. In *Proceedings of the fourth ACM conference on Recommender systems*. ACM, 257–260.

[9] Qiuxia Lu, Tianqi Chen, Weinan Zhang, Diyi Yang, and Yong Yu. 2012. Serendipitous personalized ranking for top-n recommendation. In *2012 WI-IAT*. IEEE.

[10] Tomoko Murakami, Koichiro Mori, and Ryohei Orihara. 2007. Metrics for evaluating the serendipity of recommendation lists. In *Annual conference of the Japanese society for artificial intelligence*. Springer, 40–46.

[11] Tien T Nguyen, Pik-Mai Hui, F Maxwell Harper, Loren Terveen, and Joseph A Konstan. 2014. Exploring the filter bubble: the effect of using recommender systems on content diversity. In *Proceedings of the 23rd WWW*. ACM.

[12] Eli Pariser. 2011. *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin.

[13] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2015. Recommender systems: introduction and challenges. In *Recommender systems handbook*. Springer, 1–34.

[14] Guy Shani and Asela Gunawardana. 2011. Evaluating recommendation systems. In *Recommender systems handbook*.

[15] Chuan Shi, Binbin Hu, Xin Zhao, and Philip Yu. 2018. Heterogeneous Information Network Embedding for Recommendation. *IEEE Transactions on Knowledge and Data Engineering* (2018).

[16] Yizhou Sun and Jiawei Han. 2013. Mining heterogeneous information networks: a structural analysis approach. *Acm Sigkdd Explorations Newsletter* 14, 2 (2013), 20–28.

[17] Mi Zhang and Neil Hurley. 2008. Avoiding monotony: improving the diversity of recommendation lists. In *Proceedings of the 2008 ACM conference on Recommender systems*. ACM, 123–130.

[18] Yuan Cao Zhang, Diarmuid Ó Séaghdha, Daniele Quercia, and Tamas Jambor. 2012. Auralist: introducing serendipity into music recommendation. In *Proceedings of the fifth ACM WSDM*. ACM.

[19] Zainab Zolaktaf, Reza Babanezhad, and Rachel Pottinger. 2018. A Generic Top-N Recommendation Framework For Trading-off Accuracy, Novelty, and Coverage. In *2018 IEEE 34th ICDE*. IEEE.