Towards the Next Generation of Multi-Criteria Recommender Systems

Zhe Li
Delft University of Technology
Delft, The Netherlands
Z.Li-9@tudelft.nl

ABSTRACT

This paper presents the motivation, concepts, ideas and research questions underlying a PhD research project in the domain of recommender systems, and more specifically on multi-criteria recommendation. While we build on the existing work in this direction, we aim at introducing recommendation frameworks that do not only optimize for different criteria simultaneously, but also exploit their interrelations. For this aim, we will address three multi-criteria recommendation challenges, namely multi-modal user and item modeling, package recommendation, and user-centric recommendation. For realizing these frameworks, and in particular, for learning interactions and interrelations in the criteria space, we will rely on the state-of-the-art deep learning systems, and in particular the Generative Adversarial Networks (GANs). In addition, a novel evaluation strategy for multi-criteria recommendation targeting the maximization of the user's satisfaction will also be devised.

CCS CONCEPTS

• **Information systems → Recommender systems**; • **Human-centered computing →** *User models*;

KEYWORDS

Recommender Systems; Multi-criteria; User Modeling; User-centered Recommendation

ACM Reference Format:
Zhe Li. 2018. Towards the Next Generation of Multi-Criteria Recommender Systems. In *Twelfth ACM Conference on Recommender Systems (RecSys '18), October 2–7, 2018, Vancouver, BC, Canada.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3240323.3240326

1 INTRODUCTION

Recommender systems are software tools and techniques that help the users in their decision making processes by suggesting items that are considered helpful to the users [3]. Development and deployment of recommender systems is an effective way to tackle the information overload problem in modern society, and has been a hot R&D topic in recent years.

The utility function $R(u, i)$ steering the optimization of a recommender system typically captures a single criterion (e.g. accuracy), while the suitability of a recommendation model to a practical use case usually depends on several aspects [2]. The latter is the case because the user's preferences usually cover more than one perspective. As a result, we can observe two developments that aim at maximizing the user's satisfaction with the recommendation lists [25]. First, the metrics looking beyond accuracy and including serendipity, novelty, coverage, and diversity have been drawing more attention lately [22, 52]. Second, multi-criteria and -stakeholder recommender systems have emerged as a novel and promising research area [1, 6]. This can be observed from the attempts to develop such systems in different scenarios, including online dating [29, 42], education [4], music recommendation [26] and electronic business [19].

This PhD thesis research builds on two insights derived from observing the existing work on multi-criteria recommender systems. First, there are usually only two criteria or stakeholders to be considered. To make the most suitable recommendation, we may need to consider more factors, for which efficient, effective and scalable recommendation models are needed. Second, the branches of the model optimize each for one of the criteria, but they typically do not capture interactions among them and therefore do not learn from each other.

The overall goal of this PhD project is to bring the development of the multi-criteria recommender system to a next stage by addressing the two issues mentioned above. We elaborate on this goal and our strategy to reach it in the remainder of the paper, which is organized as follows. Section 2 presents the motivation and related work underlying the PhD thesis research. Section 3 introduces the theoretical backgrounds related to our research. In Section 4, we define and explain our research questions and give an evaluation plan. Section 5 concludes the paper.

2 RELATED WORK AND MOTIVATION

Multi-criteria recommendation can be used in many application scenarios to promote the recommendation performance. In this section we illustrate this applicability and the related open issues on the examples of three scenarios: multi-modal user and item modeling, package recommendation and user-centric recommendation.

2.1 Multi-modal User and Item Modeling

User preferences can be modeled as a function of several factors, like for example, general interest and context including time, location, mood or sentiment. Similarly, items to be recommended can be represented using features from various sources. For example, in movie recommendation tasks, posters, reviews, tags, timestamps, and video contents may all be available to characterize a movie.

RecSys '18, October 2–7, 2018, Vancouver, BC, Canada

ACM ISBN 978-1-4503-5901-6/18/10.
https://doi.org/10.1145/3240323.3240326

The research in utilizing contextual features has long been popular and drawn large attention. Time and location information was first incorporated into Matrix Factorization (MF) models [23, 43] to improve the recommendation performance. To generalize context inclusion and model higher-order interactions among users, items and context, Factorization Machines (FM) [30] were proposed and widely applied in context-aware recommendations [34, 44].

More recently, the rise of deep learning has made it possible to better exploit different feature sources for user and item modeling by generating more abstract representations, e.g. Word2Vec [27], and even modality-agnostic representations, e.g. in the context of Adversarial Cross-Modal Retrieval [39]. However, up until now, there has been little research on combined learning of representations from different modalities for user and item modeling beyond the naive feature concatenation. With multi-criteria methods, information and features from different modalities can be better exploited by learning from each other. Expected to be a vector consisting of latent features, the output of the user and item modeling can be obtained by maximizing the modeling performance of all modalities (criteria) simultaneously, leading to a new User/Item2Vec paradigm.

2.2 Package Recommendation

In contrast to the multi-criteria user modeling task which deploys information from different sources and domains, in many real-world applications, users need to consider a series of items as a whole for decision making. For example, in tourism recommender systems, the products usually appear as a "travel bag" containing different items, including tourist attractions, transports, accommodations, and diets [31]. In other recommendation domains, such as restaurants (different dishes), furniture (e.g. recommending a set of desks, chairs, and sofas at the same time), and clothes (recommending a suit or clothes for a specific outfit style), a package of items is also necessary to be recommended at the same time [41]. Current state-of-the-art methods for addressing this problem can be divided into two categories, namely heuristic approaches and model-based approaches [2]. Heuristic approaches are built upon user or item similarities. The overall similarity can be obtained from either the aggregation of the similarities computed using each criterion or a value directly computed using multi-dimensional distance metrics. Model-based approaches aim to aggregate the ratings in each individual criterion to one single overall rating. Traditional machine learning and model-based collaborative filtering techniques, including Matrix Factorization (MF) [41], linear regression [21], Markov models [51], and Support Vector Machines (SVM) [20] have been successfully applied in recent years to the modeling of the multi-criteria recommendation. By utilizing the information related to different criteria, the performance can be improved compared to recommending each item individually. However, these traditional methods still face challenges. In the recommendation process, they regard each criterion as independent and the aggregation of the items is simply a matter of combining them together. Important information may therefore be lost due to the failure to detect the dependencies among different criteria. Taking the tourism recommendation as an example again, the recommendation of the diets and tourism attractions should be correlated because tourists tend to look for restaurants near the place they are in to have lunch, but this interaction cannot be modeled using only the dining history of the user and her preferences regarding tourist attractions. Cross-domain Collaborative Filtering (CDCF) provides a way to help address this problem. In CDCF, one domain is set as the target domain and the characteristics of other auxiliary domains can be transfered to the target domain. With the knowledge from different domains, CDCF can help address the cold-start problem and improve the novelty and diversity on the target domain[5]. In the tourism example, the features of the tourism attraction including its location, traffic convenience, and the estimated number of tourists can be transferred to the restaurant domain as a reference for decision making. Several ways are widely applied to make it. One is to give the weight for each auxiliary domain based on their semantic similarity with the target domain [35]. As a priori method, it is intuitive and easy to implement, but it is not sensitive to temporal dynamics and sudden changes. As an alternative, Elkahky et. al. [11] proposed a deep learning-based end-to-end knowledge transfer method. The target domain interacts with each auxiliary domain using the Matrix Factorization model. In this way, the similarities and the knowledge transfered do not have to be explicitly represented and the new target domain with transferred knowledge can be obtained without any prior knowledge. However, in package recommendation tasks, users will have transactions with items from several domains, and many of them will act as the target domain at least for some time. For each target domain, the most suitable items in it would be chosen to add to the recommendation package. However, considering the intra-group dynamics, the combination of the most suitable items may not be the best package available. Therefore, a new model that can unify different criteria would be of help to improve the recommendation performance as well as its interpretation.

2.3 User-centric Recommendation

For researchers in user-oriented fields such as recommender systems, the ultimate goal is always getting high user satisfaction. The most intuitive and accurate way to measure the user satisfaction is user study. However, due to the limitation in time and costs, it is hard to conduct a large-scale user study [33]. As an alternative, user satisfaction is simulated by the effectiveness of evaluation metrics. It is expected that a good system effectiveness can result in good user satisfaction. However, according to Said et al. [32], high user satisfaction is not necessarily guaranteed by a high system effectiveness. Urbano et al. investigated this problem within music retrieval and recommendation based on user study [36–38]. The experimental results indicate that the system effectiveness has a positive correlation with the user satisfaction for all evaluation metrics, but metrics that place more emphasis on top ranked items tend to be more pessimistic and underestimate the user satisfaction, and vice versa. As shown in Figure 1, there always exists a bias between the real user satisfaction and the system effectiveness. In order to get a real user-centric recommendation model maximizing the user's satisfaction, this bias needs to disappear. Within a multi-criteria framework, we can unify several list-wise ranking models optimized by different metrics, which have different biases towards the user satisfaction and make them learn from each other to get

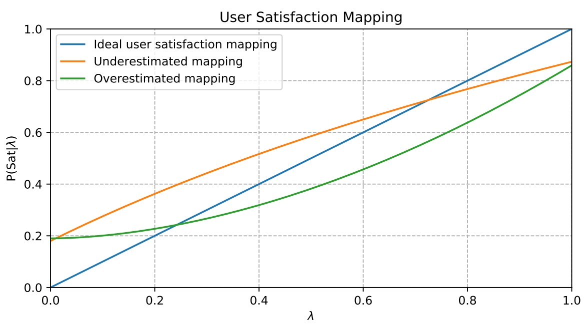

Figure 1: A brief simulation of the mapping from evaluation metrics to user satisfaction. Given one specific evaluation metric, λ is the effectiveness score the user gives to the system. $P(Sat|\lambda)$ is the probability for users to be satisfied with the system. Ideally, this mapping is a diagonal and $P(Sat|\lambda) = \lambda$, but metrics are always biased. It is worth noting that $P(Sat|0) \neq 0$ and $P(Sat|1) \neq 1$ as expected. This is because some users will have different opinions even if the system has provided ideal results in real life.

close to each other. In this way, the model is expected to converge at the point near the real user satisfaction. Thus, a user-centric recommendation model can be built without the intervention of a user study.

3 THEORETICAL BACKGROUNDS

As an application-oriented field, recommender systems do not solely rely on theories. However, the theoretical backgrounds underlying the models play an important role in obtaining good recommendation performance. In this section, we briefly present the core theories and methods related to this PhD project.

3.1 Deep Learning-based Recommender Systems

As a powerful and evolutionary technique, in recent years, deep learning has been drawing much attention in both academia and industry. Its strength in automatic learning, flexibility, and handling big data makes it get a huge success in many application scenarios including computer vision, speech recognition, and natural language processing [14]. With a large amount and diversity of data available, recommender systems are also on the way to be closely related to deep learning [47]. An increasing number of industrial recommendation services or applications are being developed based on a variety of different architectures of deep learning. For example, Google is using a wide & deep model for Apps recommendation [7]. Xtrip, a famous travel agency in China, has deployed a tourism recommender system based on autoencoders [10]. Youtube is also using deep learning techniques for video recommendation [8]. We can find that Convolutional Neural Networks (CNNs) are widely used in applications with more static features, such as visual [24] and hashtag recommendation [13], while for scenarios having a higher emphasis on sequence and semantics, such as news recommendation [28], music recommendation [16], and conversational recommender systems [49], Recurrent Neural Networks (RNNs) and Deep Structured Semantic Models (DSSMs) are also widely applied. As a novel and promising model, Generative Neural Networks (GANs) have also entered the field and have been applied in recommendation and information retrieval tasks [40, 48]. To get a better recommendation performance, different network architectures can also be used together [48].

In this PhD project, deep learning will be the core technique. We will use multiple deep learning methods for feature extraction, user and item representation, while the multi-criteria learning process will mainly be conducted using GANs. The following subsection will briefly introduce GANs and its application in recommender systems. In addition, we will also discuss the way we plan to use it for multi-criteria recommendation.

3.2 GANs for Recommender Systems

Among numerous types of deep learning models, GAN[15] is regarded as a new milestone. Basically, GAN consists of a generator G and a discriminator D. By playing a minimax game between these two parties, G tries to simulate the distribution of the real data and thus fool D, while D is also optimized to detect the fake data generated by G. The training process converges when D fails to tell the source of the images, which means the distribution of the data from G is nearly the same with the real distribution. With a strong ability in data generation and bi-lateral training, GAN shows its superiority on semi-supervised learning and the unification of two mainstream schools in machine learning, namely generative and discriminative. In addition, it can be used to find trade-offs between different criteria, as well as to let the criteria learn from each other, which is also a promising advantage in many real world applications.

As one of the most promising deep learning techniques, GANs have shown great performance on various tasks such as image captioning [9], sequence generation [12, 46], image-to-image translation [18], neural machine translation [45], and information retrieval [39, 40]. Despite the great potential shown, there is only little work focusing on the application of GANs in recommender systems (e.g. [40, 48]). [40] proposed a minimax adversarial game to unify two mainstream schools in information retrieval, which became the state-of-the-art in three different tasks, including web search, question answering and item recommendation. [48] used a similar model with a more complicated generator G combining matrix factorization and recurrent neural networks to balance the long- and short-term information in movie recommendation. However, these two methods just used a standard GAN model with only one pair of G and D. Therefore, there is only one utility function optimized in these methods, which might be insufficient to satisfy complex user needs in recommender systems.

In this project, we aim to use multiple generators, each of which is optimized based on one specific criterion. Some recent works have attempted to use multiple generators to solve this problem, such as CycleGAN [50] in image style transfer and MGAN [17] in image generation. With two pairs of Gs and Ds in the architecture, CycleGAN tries to generate fake data with two image styles simultaneously. In other words, CycleGAN tries to training two standard GANs at the same time. MGAN used several parallel generators for image generation. D is assigned with two tasks in training. On the one hand, like in other GAN models, it needs to detect the generated images; on the other hand, it needs to point out the source generator of the generated images. The training process ends when D fails both two tasks. Therefore, the generators in these models are still just learning to fit the distribution of each other, but not finding a trade-off. That is to say, for multi-criteria recommendation tasks where a trade-off among different criteria is needed, these multi-generator models cannot be applied directly because one of the criteria keeps fixed. To address this problem, we will try to make each generator regard the recommendation lists generated by other generators as the ground-truth. In the training process, they will try to fool the discriminator by getting close to each other, and the discriminator will try to distinguish between them and return the gradients or rewards for optimization. In this way, generators compete and learn from each other. The whole system is expected to converge when the discriminator shows a performance with no difference to random guess. The criteria vary according to the specific tasks. Since it can provide the interactions among different parties and use them for training, we believe that GAN can be a universal model for multi-criteria recommendation. By taking the criteria as different modalities, this model can be used for the User/Item2Vec task. Similarly, the package recommendation can be achieved by taking different needs as the criteria, and a user-centric recommendation model can be built via treating multiple evaluation metrics with different biases towards the user satisfaction as the criterion or stakeholders.

4 RESEARCH QUESTIONS

The key question underlying this PhD thesis project is how to incorporate multiple criteria in a recommendation frameworks in a way that is efficient, effective and scalable. To investigate this question from different perspectives, this overall research question is divided into the following research questions.

(**RQ1**) *How can we develop a user/item modeling method utilizing information from different modalities?* This research question is part of the multi-attribute preference modeling task. We aim to propose a User/Item2Vec embedding method based on cross-modal technologies to model each stakeholder in the recommendation process. Given meta-data from different modalities such as texts, images, and speech as criteria, we plan to use deep learning techniques for inter-modality feature engineering and then apply the cross-modal techniques to diminish the inter-modality gaps using GANs. The outcome is expected to be a well-fused multi-modal feature space, in which each user and item is represented as a fixed length vector with highly abstract latent factors.

(**RQ2**) *How can we improve the package recommendation with multi-criteria techniques?* Belonging to the type of multi-criteria preference elicitation, this research question can be regarded as the mirror question of RQ1. In RQ1, multi-criteria methods will be applied for user modeling in one specific domain with different modalities, while in RQ2 we plan to launch our research across different domains with one single data modality. The aim of this research question is to recommend a set of items strongly correlated to each other but in different domains. Criteria (domain) adaptation techniques will play a fundamental role in this task. With an end-to-end framework, the relationship among items is expected to be disclosed via the interaction among domains, users, and items. We plan to conduct our research in four recommendation domains where purchasing a set of different items is widespread, namely clothing (shoes and jewelry are also included), food and restaurants, home and kitchen, and tourism.

(**RQ3**) *Can multi-criteria techniques help build a user-centric recommender system using off-line evaluation metrics?* As a representative multi-objective optimization task, this research question intends to study the ways to build a recommendation model with high user satisfaction without the intervention of online experiments or user studies. Our aim is using multi-criteria methods to balance the bias brought by different objective evaluation metrics and thus converge to the point with a high user satisfaction. Since the user satisfaction mapping system is first proposed by Urbano et.al in the music domain [37, 38], we also start from music recommendation and then extend it to other domains. The initial design is to use GANs to build a recommendation model with multiple generators. Each generator is a list-wise ranking model optimized by one specific ranking evaluation metric. Several short term sub-questions including the selection of the metrics for optimization, the aggregation strategy of the results obtained from different generators, and the characteristics in different recommendation domains will be studied in detail under RQ4. The number of generators begins with two and will be extended to a larger value for investigation.

(**RQ4**) *How can we evaluate the performance of the multi-criteria models we propose?* This research question intends to study the evaluation of multi-criteria models. User satisfaction will be the ultimate goal for all of the models in this project. The initial design divides the evaluation as offline and online tests. In the offline part, a high user satisfaction can be demonstrated by a good performance on all kinds of evaluation metrics with different biases towards the user satisfaction. Online evaluation will be conducted based on A/B tests that can be easily controlled. All of the models proposed in this project would be evaluated with the evaluation methodology proposed in RQ4.

5 CONCLUSIONS

In this paper, we present a project about multi-criteria recommendation models that can yield better performance by criteria adaptation. As a practical result of the project, we are going to implement an experimental framework in multi-criteria recommendation that can be used on different tasks including multi-modal user modeling, package recommendation, and user-centric recommendation. Significance of our work lies in improving the quality of recommendations by detecting the relationships among different criteria and make use of them in the recommendation process.

ACKNOWLEDGMENTS

The author would like to thank Prof. Alan Hanjalic and Dr. Julián Urbano for their supervision and valuable feedback on this paper.

REFERENCES

[1] Himan Abdollahpouri, Robin Burke, and Bamshad Mobasher. 2017. Recommender Systems as Multistakeholder Environments. In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*. ACM, 347–348.

[2] Gediminas Adomavicius and YoungOk Kwon. 2015. Multi-criteria recommender systems. In *Recommender systems handbook*. Springer, 847–880.

[3] Gediminas Adomavicius and Alexander Tuzhilin. 2005. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering* 17, 6 (2005), 734–749.

[4] Robin Burke and Himan Abdollahpouri. 2016. Educational Recommendation with Multiple Stakeholders. In *Web Intelligence Workshops (WIW), IEEE/WIC/ACM International Conference on*. IEEE, 62–63.

[5] Iván Cantador, Ignacio Fernández-Tobías, Shlomo Berkovsky, and Paolo Cremonesi. 2015. Cross-domain recommender systems. In *Recommender Systems Handbook*. Springer, 919–959.

[6] Pablo Castells, Saúl Vargas, and Jun Wang. 2011. Novelty and diversity metrics for recommender systems: choice, discovery and relevance. (2011).

[7] Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, et al. 2016. Wide & deep learning for recommender systems. In *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*. ACM, 7–10.

[8] Paul Covington, Jay Adams, and Emre Sargin. 2016. Deep neural networks for youtube recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems*. ACM, 191–198.

[9] Bo Dai, Sanja Fidler, Raquel Urtasun, and Dahua Lin. 2017. Towards Diverse and Natural Image Descriptions via a Conditional GAN. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2970–2979.

[10] Xin Dong, Lei Yu, Zhonghuo Wu, Yuxia Sun, Lingfeng Yuan, and Fangxi Zhang. 2017. A Hybrid Collaborative Filtering Model with Deep Structure for Recommender Systems.. In *AAAI*. 1309–1315.

[11] Ali Mamdouh Elkahky, Yang Song, and Xiaodong He. 2015. A multi-view deep learning approach for cross domain user modeling in recommendation systems. In *Proceedings of the 24th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 278–288.

[12] William Fedus, Ian Goodfellow, and Andrew M Dai. 2018. MaskGAN: Better Text Generation via Filling in the _. *arXiv preprint arXiv:1801.07736* (2018).

[13] Yuyun Gong and Qi Zhang. 2016. Hashtag Recommendation Using Attention-Based Convolutional Neural Network.. In *IJCAI*. 2782–2788.

[14] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. 2016. *Deep learning*. Vol. 1. MIT press Cambridge.

[15] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Advances in neural information processing systems*. 2672–2680.

[16] Kartik Gupta, Noveen Sachdeva, and Vikram Pudi. 2018. Explicit Modelling of the Implicit Short Term User Preferences for Music Recommendation. In *European Conference on Information Retrieval*. Springer, 333–344.

[17] Quan Hoang, Tu Dinh Nguyen, Trung Le, and Dinh Phung. 2018. MGAN: Training generative adversarial nets with multiple generators.

[18] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. Image-To-Image Translation With Conditional Adversarial Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1125–1134.

[19] Dietmar Jannach and Gediminas Adomavicius. 2017. Price and Profit Awareness in Recommender Systems. *arXiv preprint arXiv:1707.08029* (2017).

[20] Dietmar Jannach, Zeynep Karakaya, and Fatih Gedikli. 2012. Accuracy improvements for multi-criteria recommender systems. In *Proceedings of the 13th ACM conference on electronic commerce*. ACM, 674–689.

[21] Tanisha Jhalani, Vibhor Kant, and Pragya Dwivedi. 2016. A linear regression approach to multi-criteria recommender system. In *International Conference on Data Mining and Big Data*. Springer, 235–243.

[22] Marius Kaminskas and Derek Bridge. 2017. Diversity, serendipity, novelty, and coverage: A survey and empirical analysis of beyond-accuracy objectives in recommender systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 7, 1 (2017), 2.

[23] Yehuda Koren, Robert Bell, and Chris Volinsky. 2009. Matrix factorization techniques for recommender systems. *Computer* 42, 8 (2009).

[24] Qiang Liu, Shu Wu, and Liang Wang. 2017. DeepStyle: Learning User Preferences for Visual Recommendation. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 841–844.

[25] Sean M McNee, John Riedl, and Joseph A Konstan. 2006. Being accurate is not enough: how accuracy metrics have hurt recommender systems. In *CHI'06 extended abstracts on Human factors in computing systems*. ACM, 1097–1101.

[26] Zsolt Mezei and Carsten Eickhoff. 2017. Evaluating Music Recommender Systems for Groups. *arXiv preprint arXiv:1707.09790* (2017).

[27] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. 3111–3119.

[28] Keunchan Park, Jisoo Lee, and Jaeho Choi. 2017. Deep Neural Networks for News Recommendations. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. ACM, 2255–2258.

[29] Luiz Pizzato, Tomek Rej, Thomas Chung, Irena Koprinska, and Judy Kay. 2010. RECON: a reciprocal recommender for online dating. In *Proceedings of the fourth ACM conference on Recommender systems*. ACM, 207–214.

[30] Steffen Rendle. 2010. Factorization machines. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*. IEEE, 995–1000.

[31] Francesco Ricci, Bora Arslan, Nader Mirzadeh, and Adriano Venturini. 2002. ITR: a case-based travel advisory system. In *European Conference on Case-Based Reasoning*. Springer, 613–627.

[32] Alan Said, Ben Fields, Brijnesh J Jain, and Sahin Albayrak. 2013. User-centric evaluation of a k-furthest neighbor collaborative filtering recommender algorithm. In *Proceedings of the 2013 conference on Computer supported cooperative work*. ACM, 1399–1408.

[33] Guy Shani and Asela Gunawardana. 2011. Evaluating recommendation systems. In *Recommender systems handbook*. Springer, 257–297.

[34] Yue Shi, Alexandros Karatzoglou, Linas Baltrunas, Martha Larson, Alan Hanjalic, and Nuria Oliver. 2012. TFMAP: optimizing MAP for top-n context-aware recommendation. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*. ACM, 155–164.

[35] Krishna Mohan PD Shrivastva, Shailendra Singh, et al. 2016. Cross domain recommendation using semantic similarity and tensor decomposition. *Procedia Computer Science* 85 (2016), 317–324.

[36] Julián Urbano. 2013. *Evaluation in audio music similarity*. Ph.D. Dissertation. PhD thesis, University Carlos III of Madrid.

[37] Julián Urbano, J Stephen Downie, Brian Mcfee, and Markus Schedl. 2012. How Significant is Statistically Significant? The case of Audio Music Similarity and Retrieval.. In *ISMIR*. 181–186.

[38] Julián Urbano and Mónica Marrero. 2015. How do Gain and Discount Functions Affect the Correlation between DCG and User Satisfaction?. In *European conference on information retrieval*. Springer, 197–202.

[39] Bokun Wang, Yang Yang, Xing Xu, Alan Hanjalic, and Heng Tao Shen. 2017. Adversarial Cross-Modal Retrieval. In *Proceedings of the 2017 ACM on Multimedia Conference*. ACM, 154–162.

[40] Jun Wang, Lantao Yu, Weinan Zhang, Yu Gong, Yinghui Xu, Benyou Wang, Peng Zhang, and Dell Zhang. 2017. Irgan: A minimax game for unifying generative and discriminative information retrieval models. In *Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM, 515–524.

[41] Agung Toto Wibowo, Advaith Siddharthan, Chenghua Lin, and Judith Masthoff. 2017. Matrix Factorization for Package Recommendations. In *Proceedings of the RecSys 2017 Workshop on Recommendation in Complex Scenarios (ComplexRec 2017)*. CEUR-WS.

[42] Peng Xia, Benyuan Liu, Yizhou Sun, and Cindy Chen. 2015. Reciprocal recommendation system for online dating. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*. ACM, 234–241.

[43] Liang Xiang and Qing Yang. 2009. Time-dependent models in collaborative filtering based recommender system. In *Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology-Volume 01*. IEEE Computer Society, 450–457.

[44] Jun Xiao, Hao Ye, Xiangnan He, Hanwang Zhang, Fei Wu, and Tat-Seng Chua. 2017. Attentional factorization machines: Learning the weight of feature interactions via attention networks. *arXiv preprint arXiv:1708.04617* (2017).

[45] Zhen Yang, Wei Chen, Feng Wang, and Bo Xu. 2017. Improving neural machine translation with conditional sequence generative adversarial nets. *arXiv preprint arXiv:1703.04887* (2017).

[46] Lantao Yu, Weinan Zhang, Jun Wang, and Yong Yu. 2017. SeqGAN: Sequence Generative Adversarial Nets with Policy Gradient.. In *AAAI*. 2852–2858.

[47] Shuai Zhang, Lina Yao, and Aixin Sun. 2017. Deep learning based recommender system: A survey and new perspectives. *arXiv preprint arXiv:1707.07435* (2017).

[48] Wei Zhao, Wenyou Wang, Jianbo Ye, Yongqiang Gao, Min Yang, Zhou Zhao, and Xiaojun Chen. 2017. Leveraging Long and Short-term Information in Content-aware Movie Recommendation. *arXiv preprint arXiv:1712.09059* (2017).

[49] Chunyi Zhou, Yuanyuan Jin, Kai Zhang, Jiahao Yuan, Shengyuan Li, and Xiaoling Wang. 2018. MusicRoBot: Towards Conversational Context-Aware Music Recommender System. In *International Conference on Database Systems for Advanced Applications*. Springer, 817–820.

[50] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint arXiv:1703.10593* (2017).

[51] Tao Zhu, Patrick Harrington, Junjun Li, and Lei Tang. 2014. Bundle recommendation in ecommerce. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*. ACM, 657–666.

[52] Cai-Nicolas Ziegler, Sean M McNee, Joseph A Konstan, and Georg Lausen. 2005. Improving recommendation lists through topic diversification. In *Proceedings of the 14th international conference on World Wide Web*. ACM, 22–32.