Overview of Serendipity in Recommender Systems

Denis Kotkov[0000−0003−4729−1995]

University of Helsinki
Department of Computer Science
Yliopistonkatu 4, 00100 Helsinki, Finland
kotkov.denis.ig@gmail.com

Abstract. Has it ever happened to you that services like Spotify, Netflix or YouTube showed you recommendations on the same topic over and over again? This might be caused by the lack of serendipity in recommender systems of these services. Recommender systems are software tools that suggest items, such as audio recordings or videos, of interest to users. Meanwhile, serendipity is the property of these systems, which indicates the degree, to which they suggest items that pleasantly surprise users. In this talk, I will provide an overview of serendipity in recommender systems. In particular, I will talk about how the concept of serendipity has been defined and measured in recommender systems, and what experiments have been conducted to investigate this concept. I will also touch on recommendation algorithms designed to suggest serendipitous items and discuss future directions of the topic.

Keywords: serendipity, recommender systems, overview

1 Introduction

Recommender systems are software tools that suggest items, such as movies or articles, to users [15]. To optimize for items that users would enjoy, recommender systems tend to generate recommendations that are similar to what users would usually consume and, therefore, could potentially be found without the aid of a recommender system. To overcome this problem, system designers take serendipity into account during system optimization [7]. In this talk, I will provide an overview of this topic and discuss definitions of serendipity in recommender systems, user studies conducted to investigate it, datasets containing serendipity labels, serendipity-oriented recommendation algorithms and future directions.

This preprint has not undergone peer review (when applicable) or any post-submission improvements or corrections. The Version of Record of this contribution is published in Web Engineering, ICWE 2024, Lecture Notes in Computer Science, vol 14629, and is available online at https://doi.org/10.1007/978-3-031-62362-2_43.

2 Definitions

According to the recent study [7], there are three definitions of serendipity that can be applied to recommender systems: generalized serendipity, RecSys serendipity and user serendipity. **Generalized serendipity** is based on how the term is defined in social sciences, which corresponds closely with the dictionary definition[1]: "*luck that takes the form of finding valuable or pleasant things that are not looked for*". **RecSys serendipity** is based on how the term has been defined historically in the recommender systems literature, which differs from the definition in social sciences. There is no consensus on the definition of serendipity in recommender systems [7, 5]. However, according to most authors, an item needs to correspond to one or more of the following components to be serendipitous [5]: relevance, novelty and unexpectedness. Relevance indicates that the item is beneficial to the user, while novelty that the user has limited level of familiarity with the item [5, 16]. The unexpectedness component has a number of definitions. For example, according to one of the definitions, an item is unexpected to the user if the user does not think that they would have come across this item by themselves [5, 13]. Lastly, **user serendipity** is based on each user's personal understanding of serendipity and, therefore, can cover a broad range of meanings.

3 Datasets

The following three publicly available datasets contain serendipity labels, i.e. information on whether a particular item is considered serendipitous by the user: Serendipity 2018, Taobao Serendipity and SerenLens datasets. **The Serendipity 2018 dataset** was collected in the movie recommender system MovieLens[2] [5]. The dataset contains 10 million relevance ratings, 2,150 RecSys serendipity labels on movies and movie metadata. The serendipity ratings include user ratings of statements regarding relevance, two variations of novelty and four variations of unexpectedness.

The Taobao Serendipity dataset was collected in Taobao, a popular Chinese mobile e-commerce application [20, 1]. The dataset contains 11,383 RecSys serendipity ratings on products. The authors used the following statement to measure serendipity: *"The item recommended to me is a pleasant surprise"* [1]. The dataset also contains extensive information on users, such as age, gender and previous purchases.

The SerenLens dataset is based on user reviews [4]. To generate the dataset, the authors selected a set of reviews and recruited Amazon Mechanical Turk workers to annotate them. The workers needed to specify whether the review indicated that the item was serendipitous to the review author. Overall, the SerenLens dataset contains 265,037 serendipity labels on books and 74,967 labels on movies.

[1] https://www.britannica.com/dictionary/serendipity
[2] https://movielens.org/

4 User studies

A few studies investigated serendipity in recommender systems with real users. For example, Kotkov et al. conducted a user study in MovieLens, where users retrospectively indicated if particular movies were RecSys serendipitous to them [5]. The authors investigated associations between different variations of RecSys serendipity and user behavior. The results suggested that most variations of RecSys serendipity are positively associated with preference broadening. The authors also found that ratings predicted by MovieLens, popularity, content-based and collaborative similarity to a user profile are effective predictors of whether an item is considered RecSys serendipitous by the user.

Chen et al. conducted a user study where they collected user responses to survey questions regarding RecSys serendipity in the e-commerce domain [1]. According to the results of the study, RecSys serendipity is positively associated with user satisfaction, purchase intention and timeliness.

Smets et al. carried out a survey on venues in an urban recommender system [16]. In the survey, the authors included questions regarding relevance, novelty, diversity, RecSys serendipity, satisfaction and conversion. The authors found that the more often the users visit venues, the higher the rate of them finding RecSys serendipitous venues.

Kotkov et al. ran a field study in Soulie[3], a recommender system that suggests articles to users [10]. In the study, the users were interacting with the articles and were prompted to reply to surveys. Based on user replies, the authors labeled articles RecSys, generalized and user serendipitous. The authors found that RecSys serendipity misses items that should be considered serendipitous according to generalized and user serendipity. Similarly, the authors found that user understanding of serendipity differs from generalized and RecSys serendipity. Finally, the authors discovered that different types of serendipity are associated with different patterns of user behavior.

5 Algorithms

There have been various algorithms designed to recommend serendipitous items. To achieve this goal, serendipity-oriented algorithms often rerank the output of accuracy-oriented algorithms. For example, the serendipity-oriented greedy algorithm improves serendipity of accuracy-oriented algorithms through diversification [11]. Another strategy to improve serendipity is to modify an accuracy-oriented algorithm. For example, Zheng et al. modified the objective function of PureSVD [2] to improve serendipity [21].

Due to the limited number of serendipity labels in datasets, there have been efforts to utilize transfer learning for serendipity improvement. For example, Pandey et al. trained a deep learning recommendation algorithm based on relevance ratings and tuned it based on serendipity labels to mitigate data sparsity [14].

[3] https://www.soulie.io/

6 Future directions

Future directions of serendipity in recommender systems include several key areas: contextual factors, user and item characteristics, the impact of user interfaces, and cross-domain recommendations. **Context factors,** such as time of day or weather, have demonstrated a significant influence on recommendation accuracy, implying their potential to affect serendipitous discoveries [17]. Similarly, **user and item characteristics,** such as user age or item popularity, have been linked to serendipity in recommender systems, suggesting their importance for other serendipity types [19].

User interfaces play a pivotal role in shaping user perceptions of recommendations, and therefore can have an impact on serendipity. For instance, recommendation explanations can affect user interest in suggested items [8]. Another example is MovieTuner, the system that allows users to fine-tune their recommendation preferences, such as adjusting the intensity of certain features, e.g. “more comedy” or “less mafia” [18]. To design MovieTuner, the authors used the tag genome dataset, which indicates the degree to which a particular tag applies to an item [18, 6]. Tag genome potentially enables the creation of user interfaces tailored to enhance serendipity in a single- or cross-domain settings [9].

Cross-domain recommender systems leverage data from various domains to mitigate the data sparsity problem [3]. This problem is particularly pertinent to serendipity due to the difficulty in labeling serendipitous encounters. Cross-domain approaches offer promising solutions for recommending serendipitous items by capitalizing on multiple data sources [12].

References

1. Chen, L., Yang, Y., Wang, N., Yang, K., Yuan, Q.: How serendipity improves user satisfaction with recommendations? a large-scale user evaluation. In: The world wide web conference. pp. 240–250 (2019)
2. Cremonesi, P., Koren, Y., Turrin, R.: Performance of recommender algorithms on top-n recommendation tasks. In: Proceedings of RecSys’10. pp. 39–46. ACM, New York, NY, USA (2010). https://doi.org/10.1145/1864708.1864721, http://doi.acm.org/10.1145/1864708.1864721
3. Fu, Z., Niu, X., Maher, M.L.: Deep learning models for serendipity recommendations: a survey and new perspectives. ACM Computing Surveys **56**(1), 1–26 (2023)
4. Fu, Z., Niu, X., Yu, L.: Wisdom of crowds and fine-grained learning for serendipity recommendations. In: Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 739–748 (2023)
5. Kotkov, D., Konstan, J.A., Zhao, Q., Veijalainen, J.: Investigating serendipity in recommender systems based on real user feedback. In: Proceedings of the 33rd Annual ACM Symposium on Applied Computing. pp. 1341–1350 (2018)
6. Kotkov, D., Maslov, A., Neovius, M.: Revisiting the tag relevance prediction problem. In: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. p. 1768–1772. SIGIR ’21, Association for Computing Machinery, New York, NY, USA (2021). https://doi.org/10.1145/3404835.3463019

7. Kotkov, D., Medlar, A., Glowacka, D.: Rethinking serendipity in recommender systems. In: ACM SIGIR Conference on Human Information Interaction and Retrieval. CHIIR '23, Association for Computing Machinery, New York, NY, USA (2023)
8. Kotkov, D., Medlar, A., Liu, Y., Glowacka, D.: On the negative perception of cross-domain recommendations and explanations. In: Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '24, Association for Computing Machinery, New York, NY, USA (2024). https://doi.org/10.1145/3626772.3657735
9. Kotkov, D., Medlar, A., Maslov, A., Satyal, U.R., Neovius, M., Glowacka, D.: The tag genome dataset for books. In: ACM SIGIR Conference on Human Information Interaction and Retrieval. p. 353–357. CHIIR '22, Association for Computing Machinery, New York, NY, USA (2022). https://doi.org/10.1145/3498366.3505833
10. Kotkov, D., Medlar, A., Triin, K., Glowacka, D.: The dark matter of serendipity in recommender systems. In: Proceedings of the 2024 ACM SIGIR Conference on Human Information Interaction and Retrieval. CHIIR '24, Association for Computing Machinery, New York, NY, USA (2024). https://doi.org/10.1145/3627508.3638342
11. Kotkov, D., Veijalainen, J., Wang, S.: A serendipity-oriented greedy algorithm for recommendations. In: Proceedings of the 13th International Conference on Web Information systems and Technologies. SCITEPRESS (2017)
12. Kotkov, D., Wang, S., Veijalainen, J.: Improving serendipity and accuracy in cross-domain recommender systems. In: International Conference on Web Information Systems and Technologies. pp. 105–119. Springer (2017)
13. Kotkov, D., Zhao, Q., Launis, K., Neovius, M.: Clusterexplorer: Enable user control over related recommendations via collaborative filtering and clustering. In: Proceedings of the 2020 ACM conference on Recommender systems (2020)
14. Pandey, G., Kotkov, D., Semenov, A.: Recommending serendipitous items using transfer learning. In: Proceedings of the 27th ACM international conference on information and knowledge management. pp. 1771–1774 (2018)
15. Ricci, F., Rokach, L., Shapira, B.: Recommender Systems Handbook, chap. Introduction to Recommender Systems Handbook, pp. 1–35. Springer US (2011)
16. Smets, A., Vannieuwenhuyze, J., Ballon, P.: Serendipity in the city: User evaluations of urban recommender systems. Journal of the Association for Information Science and Technology **73**(1), 19–30 (2022)
17. Trattner, C., Oberegger, A., Marinho, L., Parra, D.: Investigating the utility of the weather context for point of interest recommendations. Information Technology & Tourism **19**, 117–150 (2018)
18. Vig, J., Sen, S., Riedl, J.: The tag genome: Encoding community knowledge to support novel interaction. ACM Trans. Interact. Intell. Syst. **2**(3), 13:1–13:44 (Sep 2012). https://doi.org/10.1145/2362394.2362395
19. Wang, N., Chen, L.: How do item features and user characteristics affect users' perceptions of recommendation serendipity? a cross-domain analysis. User Modeling and User-Adapted Interaction pp. 1–39 (2022)
20. Wang, N., Chen, L., Yang, Y.: The impacts of item features and user characteristics on users' perceived serendipity of recommendations. In: Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization. pp. 266–274 (2020)
21. Zheng, Q., Chan, C.K., Ip, H.H.: An unexpectedness-augmented utility model for making serendipitous recommendation. In: Advances in Data Mining: Applications and Theoretical Aspects, vol. 9165, pp. 216–230. Springer International Publishing (2015)