This is the final draft of the following article:
Denis Kotkov, Jari Veijalainen, and Shuaiqiang Wang. Challenges of serendipity in recommender systems. In Proceedings of the 12th International Conference on Web Information Systems and Technologies, volume 2, pages 251-256. SCITEPRESS, 2016. doi: 10.5220/0005879802510256.

Challenges of Serendipity in Recommender Systems

Denis Kotkov, Jari Veijalainen and Shuaiqiang Wang
University of Jyvaskyla, Dept. of Computer Science and Information Systems,
P.O.Box 35, FI-40014 University of Jyvaskyla, Jyvaskyla, Finland
deigkotk@student.jyu.fi, {jari.veijalainen, shuaiqiang.wang}@jyu.fi

Keywords: Relevance, Unexpectedness, Novelty, Serendipity, Recommender Systems, Evaluation Metrics, Challenges

Abstract: Most recommender systems suggest items similar to a user profile, which results in boring recommendations limited by user preferences indicated in the system. To overcome this problem, recommender systems should suggest serendipitous items, which is a challenging task, as it is unclear what makes items serendipitous to a user and how to measure serendipity. The concept is difficult to investigate, as serendipity includes an emotional dimension and serendipitous encounters are very rare. In this paper, we discuss mentioned challenges, review definitions of serendipity and serendipity-oriented evaluation metrics. The goal of the paper is to guide and inspire future efforts on serendipity in recommender systems.

1 INTRODUCTION

With the growth of information on the Internet it becomes difficult to find content interesting to a user. Hopefully, recommender systems are designed to solve this problem. In this paper, the term *recommender system* refers to a software tool that suggests items of use to users (Ricci et al., 2011). An item is a piece of information that refers to a tangible or digital object, such as a good, a service or a process that a recommender system suggests to the user in an interaction through the Web, email or text message (Ricci et al., 2011). For example, an item can be a reference to a movie, a song or even a friend in an online social network.

Recommender systems and search engines are different kinds of systems that aim at satisfying user information needs. Traditionally, a search engine receives a query and, in some cases, a user profile as an input and provides a set of the most suitable items in response (Smyth et al., 2011). In contrast, a recommender system does not receive any query, but a user profile and returns a set of items users would enjoy (Ricci et al., 2011). The term *user profile* refers to actions a user performed with items in the past. A user profile is often represented by ratings a user gave to items.

Recommender systems are widely adopted by different services to increase turnover (Ricci et al., 2011). Meanwhile, users need a recommender system to discover novel and interesting items, as it is demanding to search items manually among the overwhelming number of them (Shani and Gunawardana, 2011; Celma Herrada, 2009).

Most recommendation algorithms are evaluated based on accuracy that indicates how good an algorithm is at offering interesting items regardless of how obvious and familiar to a user the suggestions are (de Gemmis et al., 2015).To achieve high accuracy, recommender systems tend to suggest items similar to a user profile (Tacchini, 2012). As a result, the user receives recommendations only of items similar to items the user rated initially. Accuracy-based algorithms limit the number of items that can be recommended to the user (so-called *overspecialization* problem), which lowers user satisfaction (Celma Herrada, 2009; Tacchini, 2012). To overcome overspecialization problem and broaden user preferences, a recommender system should suggest serendipitous items.

Suggesting serendipitous items is challenging (Foster and Ford, 2003). Currently, there is no consensus on definition of serendipity in recommender systems (Maksai et al., 2015; Iaquinta et al., 2010). It is difficult to investigate serendipity, as the concept includes an emotional dimension (Foster and Ford, 2003) and serendipitous encounters are very rare (André et al., 2009). As different definitions of serendipity have been proposed (Maksai et al., 2015; Iaquinta et al., 2010), it is not clear how to measure serendipity in recommender systems (Murakami et al., 2008; Zhang et al., 2012).

In this paper we are going to discuss mentioned challenges to guide and inspire future efforts on serendipity in recommender systems. We review definitions of serendipity. We also review and classify evaluation metrics to measure serendipity and indicate their advantages and disadvantages.

2 CHALLENGES OF SERENDIPITY IN RECOMMENDER SYSTEMS

Suggesting serendipitous items involves certain challenges. We are going to present the most important of them. Designing a serendipity-oriented recommendation algorithm requires to choose suitable objectives. It is therefore necessary to investigate how to assess serendipity in recommender systems, which requires a definition of the concept.

2.1 Definition

It is challenging to define what serendipity is in recommender systems, what kind of items are serendipitous and why, since serendipity is a complex concept (Maksai et al., 2015; Iaquinta et al., 2010).

According to the dictionary[1], serendipity is "the faculty of making fortunate discoveries by accident". The term was coined by Horace Walpole in the letter to Sir Horace Mann in 1754. The author described his unexpected discovery by referencing the fairy tale, "The Three Princes of Serendip". Horace Walpole in his letter explained that the princes were "always making discoveries, by accidents and sagacity, of things which they were not in quest of" (Remer, 1965).

One of the examples of serendipity is the discovery of penicillin. On September 3, 1928, Alexander Fleming was sorting petri dishes and noticed a dish with a blob of mold. The mold in the dish killed one type of bacteria, but did not affect another. Later, the active substance from the mold was named penicillin and used to treat a wide range of diseases, such as pneumonia, skin infections or rheumatic fever. The discovery of penicillin can be regarded as serendipitous, as it led to the result positive for the researcher and happened by accident.

To introduce discoveries similar to the discovery of penicillin in recommender systems, it is necessary to define and strictly formalize the concept of serendipity. We therefore reviewed definitions employed in publications on recommender systems.

[1] http://www.thefreedictionary.com/serendipity

Corneli et al. investigated serendipity in a computational context including recommender systems and proposed the framework to describe the concept (Corneli et al., 2014). The authors considered an essential key condition, *focus shift*. A focus shift happens when something that initially was uninteresting, neutral or even negative becomes interesting.

One of definitions used in recommender systems was employed by Zhang et al.: "Serendipity represents the "unusualness" or "surprise" of recommendations" (Zhang et al., 2012). The definition does not require serendipitous items to be interesting to a user, but surprising.

In contrast, Maksai et al. indicated that serendipitous items must be not only unexpected (surprising), but also useful to a user: "Serendipity is the quality of being both unexpected and useful" (Maksai et al., 2015).

Adamopoulos and Tuzhilin used another definition. The authors mentioned the following components related to serendipity: unexpectedness, novelty and a positive emotional response, which can be regarded as relevance of an item for a user:

> *Serendipity, the most closely related concept to unexpectedness, involves a positive emotional response of the user about a previously unknown (novel) [...] serendipitous recommendations are by definition also novel.* (Adamopoulos and Tuzhilin, 2014).

A similar definition was employed by Iaquinta et al. According to (Iaquinta et al., 2010), serendipitous items are interesting, unexpected and novel to a user:

> *A serendipitous recommendation helps the user to find a surprisingly interesting item that she might not have otherwise discovered (or it would have been really hard to discover). [...] Serendipity cannot happen if the user already knows what is recommended to her, because a serendipitous happening is by definition something new. Thus the lower is the probability that user knows an item, the higher is the probability that a specific item could result in a serendipitous recommendation* (Iaquinta et al., 2010).

Definitions used in (Iaquinta et al., 2010) and (Adamopoulos and Tuzhilin, 2014) seem to correspond to the dictionary definition and the framework proposed by Corneli et al. As a serendipitous item is novel and unexpected, the item can be perceived as uninteresting, at first sight, but eventually the item will be regarded as interesting, which creates a focus shift, a necessary condition for serendipity (Corneli et al., 2014). The definition also corresponds to the dictionary definition, as novel, unexpected and interesting to a user item is likely to be a "fortunate discovery".

Publications dedicated to serendipity in recommender systems do not often elaborate the components of serendipity (Iaquinta et al., 2010; Maksai et al., 2015; Zhang et al., 2012). It is not entirely clear in what sense items should be novel and unexpected to a user.

Kapoor et al. indicated three different definitions of novelty in recommender systems (Kapoor et al., 2015):

1. *Novel to a recommender system item.* A recently added item that users have not yet assessed.
2. *Forgotten item.* A user might forget that she consumed the item some time ago in the past.
3. *Unknown item.* A user has never seen or heard about the item in her life.

In addition, Shani and Gunawardana suggested that we may regard a novel item as one not rated by the target user regardless of whether she is familiar with the item (Shani and Gunawardana, 2011).

Unexpectedness might also have different meanings depending on expectations of a user. A user might expect a recommender system to suggest items similar to her profile, popular among other users or both similar and popular (Kaminskas and Bridge, 2014; Zheng et al., 2015).

2.1.1 Serendipity in a context

Most recommender systems do not consider any contextual information, such as time, location or mood of a user (Adomavicius and Tuzhilin, 2011). Meanwhile, the context may significantly affect the relevance of items for a user (Adomavicius and Tuzhilin, 2011). An item that was relevant for a user yesterday might not be relevant tomorrow. A context may include any information related to recommendations. For example, a recommender system may consider current weather to suggest a place to visit. Context-aware recommender systems use contextual information to suggest items interesting to a user.

Serendipity depends on a context, as each of its components is context-dependant. An item that was relevant, novel and unexpected to a user in one context might not be perceived the same in another context. The inclusion of a context affects the definition of serendipity. For example, contextual unexpectedness would indicate how unexpected an item is in a given context, which might be different from unexpectedness in general. Serendipity might consist of novelty, unexpectedness and relevance in a given context.

As the context has a very broad definition (Dey, 2001), it is challenging to estimate what contextual information is the most important in a particular situation. For example, weather is an important factor for most outdoor activities, while user mood is important for music suggestion (Kaminskas and Ricci, 2012).

2.1.2 Serendipity in cross-domain recommender systems

Most recommender systems suggest items from a single domain, where the term domain refers to "a set of items that share certain characteristics that are exploited by a particular recommender system" (Fernández-Tobías et al., 2012). These characteristics are items' attributes and users' ratings. Different domains can be represented by movies and books, songs and places, MovieLens[2] movies and Netflix[3] movies (Cantador and Cremonesi, 2014).

Recommender systems that suggest items using multiple domains are called cross-domain recommender systems. Cross-domain recommender systems can use information from several domains, suggest items from different domains or both consider different domains and suggest items from them (Cantador and Cremonesi, 2014). For example, a cross-domain recommender system may take into account movie preferences of a user and places that the user visits to recommend watching a particular movie in a cinema suitable for the user.

Consideration of additional domains affects the definition of serendipity, as cross-domain recommender systems may suggest combinations of items. It is questionable whether in this case items in the recommended combination must be novel and unexpected.

2.1.3 Discussion

According to literature review, to date, there is no consensus on definition of serendipity in recommender systems (Maksai et al., 2015; Iaquinta et al., 2010). We suggest that the definition of serendipity should include combinations of items from different domains and a context, which might encourage researchers to propose serendipity-oriented recommendation algorithm that would be more satisfying to users. For example, suppose, two young travelers walk in a cold rain in a foreign city without much money. A recommendation of a hostel would be obvious in this situation, as the travelers would look for a hostel on their own. A suggestion of sleeping in a local cinema,

[2]https://movielens.org/

[3]https://www.netflix.com/

Table 1: Notation

$I = (i_1, i_2, ..., i_n)$	the set of items
$F = (f_1, f_2, ..., f_z)$	feature set
$i = (f_{i,1}, f_{i,2}, ..., f_{i,z})$	representation of item i
$U = (u_1, u_2, ..., u_n)$	the set of users
$I_u, I_u \subseteq I$	the set of items rated by user u (user profile)
$R_u, R_u \subseteq I$	the set of items recommended to user u
$rel_u(i)$	1 if item i relevant for user u and 0 otherwise

which would cost less than a hostel, is likely to be serendipitous in that situation. The recommendation would be even more satisfying to the travelers if the recommender system also suggested the longest and cheapest movie in that cinema and an energetic song to cheer up the travelers.

2.2 Emotional dimension

Relevance of an item for a user might depend on user mood (Kaminskas and Ricci, 2012). This contextual information is difficult to capture without explicitly asking the user. As serendipity is a complex concept, which includes relevance (Iaquinta et al., 2010; Adamopoulos and Tuzhilin, 2014), this concept depends on the current user mood in a higher degree. An emotional dimension makes serendipity unstable and therefore difficult to investigate (Foster and Ford, 2003).

2.3 Lack of serendipitous encounters

As serendipitous items must be relevant, novel and unexpected to a user, they are rare (André et al., 2009) and valuable. Due to the lack of observations it is difficult to make assumptions regarding serendipity that would be reasonable in most cases.

2.4 Evaluation metrics

We are going to review evaluation metrics that measure serendipity in recommender systems. As different metrics have been proposed (Murakami et al., 2008; Kaminskas and Bridge, 2014; Zhang et al., 2012), the section provides their comparison, including advantages and disadvantages. To review evaluation metrics, we first present notation in table 1.

The following evaluation metrics consider a recommender system with I available items and U users. User u rates or interacts with items $I_u, I_u \subseteq I$. A recommender system suggests R_u items to user u. Each item $i, i \in I$ is represented as a vector $i = (f_{i,1}, f_{i,2}, ..., f_{i,z})$ in a multidimensional feature space F. For example, a feature can be a genre of a movie on a web-site. If $F = (drama, crime, action)$ then the movie "*The Shawshank Redemption*" can be represented as $i_{Shawshank} = (0.4, 0.4, 0.1)$.

Seeking to measure serendipity of a recommender system, researchers proposed different evaluation metrics. Based on reviewed literature we classify them into three categories: content-based unexpectedness, collaborative unexpectedness and primitive recommender-based serendipity.

2.4.1 Content-based unexpectedness

Content-based unexpectedness metrics are based on attributes of items. These metrics indicate the dissimilarity of suggestions to a user profile.

One of the content-based unexpectedness metrics was proposed by Vargas and Castells (Vargas and Castells, 2011). Later, the metric was adopted by Kaminskas and Bridge to measure unexpectedness (Kaminskas and Bridge, 2014). The authors suggested that serendipity consists of two components: relevance and unexpectedness. Content-based unexpectedness metrics can be used to measure unexpectedness, while accuracy metrics such as root mean square error (RMSE), mean absolute error (MAE) or precision (Ekstrand et al., 2011) can be used to assess relevance. The metric is calculated as follows:

$$unexp_c(i,u) = \frac{1}{|I_u|} \sum_{j \in I_u} 1 - sim(i,j) \quad (1)$$

where $sim(i,j)$ is any kind of similarity between items i and j. For example, it might be content-based cosine distance (Lops et al., 2011).

2.4.2 Collaborative unexpectedness

Collaborative unexpectedness metrics are based on ratings users gave to items. Kaminskas and Bridge proposed a metric that can measure unexpectedness based on user ratings (Kaminskas and Bridge, 2014). User ratings can indicate similarities between items. Items can be considered similar if they are rated by the same set of users. The authors therefore proposed a co-occurrence unexpectedness metric, which is based on normalized point-wise mutual information:

$$unexp_r(i,u) = \frac{1}{|I_u|} \sum_{j \in I_u} -\log_2 \frac{p(i,j)}{p(i)p(j)} / \log_2 p(i,j) \quad (2)$$

where $p(i)$ is the probability that users have rated item i, while $p(i,j)$ is the probability that the same users have rated items i and j.

2.4.3 Primitive recommender-based serendipity

The metric is based on suggestions generated by a primitive recommender system, which is expected to generate recommendations with low unexpectedness. Originally the metric was proposed by Murakami (Murakami et al., 2008) and later modified (Ge et al., 2010; Adamopoulos and Tuzhilin, 2014). Modification proposed by Adamopoulos and Tuzhilin is calculated as follows:

$$ser_{pm}(u) = \frac{1}{|R_u|} \sum_{i \in (R_u \setminus (E_u \cup PM))} rel_u(i), \qquad (3)$$

where PM is a set of items generated by a primitive recommender system, while E_u is a set items that matches interests of user u. In the experiments conducted by Adamopoulos and Tuzhilin, the primitive recommender system generated non-personalized recommendations consisting of popular and highly rated items. Meanwhile, E_u contained items similar to what user u consumes.

2.4.4 Analysis of the evaluation metrics

Content-based and collaborative metrics capture the difference between recommended items and a user profile, but have a disadvantage. These metrics measure unexpectedness separately from relevance. The high score of both metrics can be obtained by suggesting many unexpected irrelevant and expected relevant items that would probably not be serendipitous.

Depending on a primitive recommender system, the metrics based on a primitive recommender system capture item popularity and dissimilarity to a user profile, but also have a disadvantage. Primitive recommender-based metrics are sensitive to a primitive recommender system (Kaminskas and Bridge, 2014). By changing this parameter, one might obtain contradictory results.

Designing a serendipity-oriented algorithm that takes into account a context and combinations of items from different domains requires a corresponding serendipity definition and serendipity metric. An item might be represented by a combination of items from different domains and considered serendipitous, depending on a particular situation. The reviewed metrics disregard a context and additional domains due to the lack of serendipity definitions that consider this information. One of the reasons might be that recommender systems do not usually have the information on the context. Another reason might be the disadvantages of offline evaluation.

Even offline evaluation of only relevance without considering the context or additional domains may not correspond to results of experiments involving real users (Said et al., 2013; Garcin et al., 2014). Offline evaluation may help choose candidate algorithms (Shani and Gunawardana, 2011), but online evaluation is still necessary, especially in assessing serendipity, as serendipitous items are novel by definition (Iaquinta et al., 2010; Adamopoulos and Tuzhilin, 2014) and it is difficult to assess whether a user is familiar with an item without asking her.

3 CONCLUSIONS AND FUTURE RESEARCH

In this paper, we discussed challenges of serendipity in recommender systems. Serendipity is challenging to investigate, as it includes an emotional dimension, which is difficult to capture, and serendipitous encounters are very rare, since serendipity is a complex concept that includes other concepts.

According to the reviewed literature, currently there is no consensus on definition of serendipity in recommender systems, which makes it difficult to measure the concept. The reviewed serendipity evaluation metrics can be divided into three categories: content-based unexpectedness, collaborative unexpectedness and primitive recommender-based serendipity. The main disadvantage of content-based and collaborative unexpectedness metrics is that they measure unexpectedness separately from relevance, which might cause mistakes. The main disadvantage of primitive recommender-based serendipity metrics is that they are sensitive to a primitive recommender.

In our future work, we are going to propose a definition of serendipity in recommender systems, develop serendipity metrics and design recommendation algorithms that suggest serendipitous items. We are also planning to conduct experiments using precollected datasets and involving real users. We hope that this paper will guide and inspire future research on recommendation algorithms focused on user satisfaction.

4 ACKNOWLEDGEMENT

The research at the University of Jyväskylä was performed in the MineSocMed project, partially supported by the Academy of Finland, grant #268078.

REFERENCES

Adamopoulos, P. and Tuzhilin, A. (2014). On unexpectedness in recommender systems: Or how to better expect the unexpected. *ACM Transactions on Intelligent Systems and Technology*, 5(4):54:1–54:32.

Adomavicius, G. and Tuzhilin, A. (2011). Context-aware recommender systems. In *Recommender Systems Handbook*, pages 217–253. Springer US.

André, P., schraefel, m., Teevan, J., and Dumais, S. T. (2009). Discovery is never by chance: Designing for (un)serendipity. In *Proceedings of the Seventh ACM Conference on Creativity and Cognition*, pages 305–314, New York, NY, USA. ACM.

Cantador, I. and Cremonesi, P. (2014). Tutorial on cross-domain recommender systems. In *Proceedings of the 8th ACM Conference on Recommender Systems*, pages 401–402, New York, NY, USA. ACM.

Celma Herrada, Ò. (2009). *Music recommendation and discovery in the long tail*. PhD thesis, Universitat Pompeu Fabra.

Corneli, J., Pease, A., Colton, S., Jordanous, A., and Guckelsberger, C. (2014). Modelling serendipity in a computational context. *CoRR*, abs/1411.0440.

de Gemmis, M., Lops, P., Semeraro, G., and Musto, C. (2015). An investigation on the serendipity problem in recommender systems. *Information Processing & Management*, 51(5):695 – 717.

Dey, A. K. (2001). Understanding and using context. *Personal Ubiquitous Comput.*, 5(1):4–7.

Ekstrand, M. D., Riedl, J. T., and Konstan, J. A. (2011). Collaborative filtering recommender systems. *Found. Trends Hum.-Comput. Interact.*, 4(2):81–173.

Fernández-Tobías, I., Cantador, I., Kaminskas, M., and Ricci, F. (2012). Cross-domain recommender systems: A survey of the state of the art. In *Spanish Conference on Information Retrieval*.

Foster, A. and Ford, N. (2003). Serendipity and information seeking: an empirical study. *Journal of Documentation*, 59(3):321–340.

Garcin, F., Faltings, B., Donatsch, O., Alazzawi, A., Bruttin, C., and Huber, A. (2014). Offline and online evaluation of news recommender systems at swissinfo.ch. In *Proceedings of the 8th ACM Conference on Recommender Systems*, pages 169–176, New York, NY, USA. ACM.

Ge, M., Delgado-Battenfeld, C., and Jannach, D. (2010). Beyond accuracy: Evaluating recommender systems by coverage and serendipity. In *Proceedings of the Fourth ACM Conference on Recommender Systems*, pages 257–260, New York, NY, USA. ACM.

Iaquinta, L., Semeraro, G., de Gemmis, M., Lops, P., and Molino, P. (2010). *Can a recommender system induce serendipitous encounters?* INTECH Open Access Publisher.

Kaminskas, M. and Bridge, D. (2014). Measuring surprise in recommender systems. In *Workshop on Recommender Systems Evaluation: Dimensions and Design*.

Kaminskas, M. and Ricci, F. (2012). Contextual music information retrieval and recommendation: State of the art and challenges. *Computer Science Review*, 6(23):89 – 119.

Kapoor, K., Kumar, V., Terveen, L., Konstan, J. A., and Schrater, P. (2015). "i like to explore sometimes": Adapting to dynamic user novelty preferences. In *Proceedings of the 9th ACM Conference on Recommender Systems*, pages 19–26, New York, NY, USA. ACM.

Lops, P., de Gemmis, M., and Semeraro, G. (2011). Content-based recommender systems: State of the art and trends. In *Recommender Systems Handbook*, pages 73–105. Springer US.

Maksai, A., Garcin, F., and Faltings, B. (2015). Predicting online performance of news recommender systems through richer evaluation metrics. In *Proceedings of the 9th ACM Conference on Recommender Systems*, pages 179–186, New York, NY, USA. ACM.

Murakami, T., Mori, K., and Orihara, R. (2008). Metrics for evaluating the serendipity of recommendation lists. In *New Frontiers in Artificial Intelligence*, volume 4914 of *Lecture Notes in Computer Science*, pages 40–46. Springer Berlin Heidelberg.

Remer, T. G. (1965). *Serendipity and the three princes: From the Peregrinaggio of 1557*, page 20. Norman, U. Oklahoma P.

Ricci, F., Rokach, L., and Shapira, B. (2011). *Introduction to Recommender Systems Handbook*. Springer US.

Said, A., Fields, B., Jain, B. J., and Albayrak, S. (2013). User-centric evaluation of a k-furthest neighbor collaborative filtering recommender algorithm. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, pages 1399–1408, New York, NY, USA. ACM.

Shani, G. and Gunawardana, A. (2011). Evaluating recommendation systems. In *Recommender Systems Handbook*, pages 257–297. Springer US.

Smyth, B., Coyle, M., and Briggs, P. (2011). Communities, collaboration, and recommender systems in personalized web search. In *Recommender Systems Handbook*, pages 579–614. Springer US.

Tacchini, E. (2012). *Serendipitous mentorship in music recommender systems*. PhD thesis, Ph. D. thesis., Computer Science Ph. D. School–Università degli Studi di Milano.

Vargas, S. and Castells, P. (2011). Rank and relevance in novelty and diversity metrics for recommender systems. In *Proceedings of the Fifth ACM Conference on Recommender Systems*, pages 109–116, New York, NY, USA. ACM.

Zhang, Y. C., Séaghdha, D. O., Quercia, D., and Jambor, T. (2012). Auralist: Introducing serendipity into music recommendation. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, pages 13–22, New York, NY, USA. ACM.

Zheng, Q., Chan, C.-K., and Ip, H. (2015). An unexpectedness-augmented utility model for making serendipitous recommendation. In *Advances in Data Mining: Applications and Theoretical Aspects*, volume 9165 of *Lecture Notes in Computer Science*, pages 216–230. Springer International Publishing.