

# How Do Item Features and User Characteristics Affect Users' Perceptions of Recommendation Serendipity? A Cross-Domain Analysis

Ningxia Wang (✉ [nxwang@comp.hkbu.edu.hk](mailto:nxwang@comp.hkbu.edu.hk))

Hong Kong Baptist University

Li Chen

Hong Kong Baptist University

---

## Research Article

**Keywords:** Recommender systems, serendipity, item features, user personality, curiosity, cross-domain

**Posted Date:** April 8th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1520610/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# How Do Item Features and User Characteristics Affect Users' Perceptions of Recommendation Serendipity? A Cross-Domain Analysis

Ningxia Wang<sup>1\*</sup> and Li Chen<sup>2</sup>

<sup>1\*</sup>Department of Computer Science, Hong Kong Baptist University, 34 Renfrew Rd, Kowloon Tsai, Hong Kong, 999077, Hong Kong, China.

<sup>2</sup>Department of Computer Science, Hong Kong Baptist University, 34 Renfrew Rd, Kowloon Tsai, Hong Kong, 999077, Hong Kong, China.

\*Corresponding author(s). E-mail(s): [nxwang@comp.hkbu.edu.hk](mailto:nxwang@comp.hkbu.edu.hk);  
Contributing authors: [lichen@comp.hkbu.edu.hk](mailto:lichen@comp.hkbu.edu.hk);

## Abstract

Serendipity is one of beyond-accuracy objectives for recommender systems (RSs), which aims to achieve both *relevance* and *unexpectedness* of recommendations, so as to potentially address the “filter bubble” issue of traditional accuracy-oriented RSs. However, so far most of the serendipity-oriented studies have focused on developing algorithms to consider various types of item features or user characteristics, but are largely based on their own assumptions. Few have stood from users' perspective to identify the effects of these features on users' perceptions of the serendipity of the recommendation. Therefore, in this paper, we have analyzed their effects with two user survey datasets. These are the Movielens Serendipity dataset of 467 users' responses to a retrospective survey of

their perceptions of the recommended movie's serendipity, and the Taobao Serendipity dataset of 11,383 users' perceptions of the serendipity of a recommendation received at a mobile e-commerce platform. In both datasets, we have analyzed the correlations between users' serendipity perceptions and various types of item features (i.e., *item-driven* such as popularity, *profile-driven* such as in-profile diversity, and *interaction-driven* including category-level and item-level features), as well as the influence of several user characteristics (including the Big-Five personality traits and curiosity). The results disclose both *domain-independent* and *domain-specific* observations, which may be constructive in enhancing current serendipity-oriented recommender systems by better utilizing item features and user data.

**Keywords:** Recommender systems, serendipity, item features, user personality, curiosity, cross-domain

## Self-Accessment

While many studies have aimed to propose more personalized recommendation approaches by considering the beyond-accuracy objective recommendation serendipity, we have not come across a paper that investigated how item features and user characteristics affect users' perceptions of recommendation serendipity. In this work, we analyzed their effects in two different domains (movie recommendation vs. e-commerce recommendation). The results disclose both domain-independent and domain-specific observations, which may be constructive in enhancing current serendipity-oriented recommender systems by better utilizing item features and user data.

We confirm that the novelty and topic of this work are relevant to the scope of your journal and will be of interest to its readership. As an extension of our preciously published conference paper, titled 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 (UMAP '20), this manuscript has not been published or presented elsewhere in entirety and is not under consideration by another journal. There are no conflicts of interest to declare.

## Acknowledgements

This work was supported by Hong Kong Research Grants Council (RGC) (project RGC/HKBU12201620). We are also thankful for Yonghua Yang, Keping Yang, and Quan Yuan who helped collect the data in the previous work (Wang et al., 2020). Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the collaborators and sponsor.

## 1 Introduction

Recommender systems (RSs) serve as popular tools in the current online environment to filter massive information and help users make effective decisions. Traditional RSs (e.g., collaborative filtering-based and content-based RSs) mainly aim to maximize the accuracy of item prediction, but have the potential to create the “filter bubble” problem (Matt et al., 2014; McNee et al., 2006), as recommendations that are too similar to users’ previous preferences may stop them from exploring different and new items. Because of this limitation, *serendipity-oriented RSs* have been proposed, targeted to achieve both *relevance* and *unexpectedness* (also called *surprise*) (Castells et al., 2011; Kaminskas and Bridge, 2016). A recent study has shown that increased user perception of recommendation serendipity can significantly lead to higher user satisfaction with the recommendation (Chen et al., 2019b).

4 *Item Features and User Characteristics with Recommendation Serendipity*

The idea behind the existing serendipity-oriented RSs is to capture item features that may satisfy users' needs for *pleasant unexpectedness*. For example, it was assumed that unpopular items would bring more uncertainty and hence be more unexpected to users, so the importance of unpopular items was increased when generating serendipitous recommendations (Lu et al., 2012a). In other work it was assumed that the more dissimilar an item is to the user's profile (i.e., the previously visited items), the more unexpected it would be to the user (Kotkov et al., 2018b; Zheng et al., 2015). Users' personal characteristics have also been considered in some work. For instance, highly *curious* users were assumed to be more likely to accept novel items, while users with less curiosity might prefer to receive recommendations similar to their previous preferences (Menk et al., 2017; Maccatrozzo et al., 2017; Niu and Abbas, 2017).

However, to the best of our knowledge, there has been little research looking at what item features and/or user characteristics might affect users' perceived serendipity of the recommendation. Given that serendipity involves users' positive emotional responses (i.e., pleasant unexpected feeling), the lack of validation with real users may impair the applicability of existing serendipity-oriented RSs in real-life situations (Kotkov et al., 2018a). Furthermore, there has been little discussion of the application domain of the proposed hypotheses, ignoring the differences brought by domain-specific properties. For example, in the context of e-commerce that induces certain financial risks, people may be more cautious when making decisions. Their propensity towards serendipity might therefore differ from the evaluation of a low-risk product (such as a movie recommendation).

Therefore, to fill the vacancy in the state-of-the-art research, we have focused on the following three major research questions in this work:

*RQ1:* What item features can be correlated with users' perceived serendipity of a recommendation?

*RQ2:* What user characteristics can affect their perceived serendipity, and what is the relative importance of all the considered features (i.e., including item features and user characteristics)?

*RQ3:* Are there any domain-specific differences in the above effects?

To address these three questions, we conducted analyses on two user survey datasets. One is the publicly available MovieLens Serendipity Dataset (Kotkov et al., 2018a), containing 467 users' feedback on serendipity-related objectives of movie recommendations. The other dataset is the Taobao Serendipity Dataset (Chen et al., 2019b; Wang et al., 2020), collected via a large-scale user survey on Mobile Taobao (a popular mobile e-commerce app in China). Specifically, it contains the feedback of 11,383 customers on the serendipity of recommendations they received through the app. Moreover, the dataset has the demographic information of these users (i.e., age and gender) and their psychological profiles (i.e., the Big-Five personality traits and curiosity), so that it is possible to identify their effects on users' serendipity perceptions.

Through statistical approaches, we identified several *domain-independent* and *domain-specific* observations. In particular, in both the MovieLens and Taobao datasets, items with lower popularity or higher frequencies of temporal interactions were more likely to be perceived as serendipitous by users. Regarding domain-specific observations, we found that in movie recommendations the more diverse the item categories in user profile (i.e., *in-profile diversity*), the more likely the user felt a recommendation to be serendipitous. Movies of categories that the user has less interacted with were more likely to be perceived as serendipitous. However, the results were

opposite in the Taobao dataset. Regarding user characteristics, we found that users' *curiosity* level and *age* were most likely to affect their perceptions of recommendation serendipity.

There are three major contributions of this work:

- 1). We tested the impacts of a number of item features and user characteristics on users' perceived recommendation serendipity, in two user survey datasets of different domains (i.e., movies and e-commerce).
- 2). We measured the relative importance of different features as well as the influence of profile length on the correlations between item features and users' perceived serendipity.
- 3). We identified differences in users' serendipity perceptions of movie and e-commerce recommendations, which could be constructive in developing dedicated serendipity-oriented recommender systems that consider domain specialities.

The remainder is organized as follows. We first introduce the related work on serendipity-oriented RSs, with a summary of item features and user characteristics considered (Section 2). We then present the descriptive statistics of the two user survey datasets used in Section 3, followed by investigated item features in Section 4, user characteristics in Section 5, and the results analysis in Section 6. Subsequently, we answer the four research questions in Section 7, discuss the implications and limitations of our work in Section 7.3, and conclude the work in Section 8.

## 2 Background

The original definition of serendipity is “[...] *making discoveries, by accidents and sagacity, of things which they were not in quest for [...]*” (Walpole, 1960; Bogers and Björneborn, 2013; Maccatrozzo et al., 2017). In recommender

systems (RSs), it has been considered for breaking through barriers, especially the “filter bubble”, of traditional accuracy-oriented approaches. In [Kotkov et al. \(2016\)](#), serendipity-oriented recommendation methods were classified into three categories: re-ranking (post-processing after conventional recommendation algorithms), modification (taking into account new factors to change the objective functions of conventional recommendation systems), and new algorithms (from the framework of conventional recommender systems). In this section we introduce related methods from the aspects of **item features** and **user characteristics** they have considered.

## 2.1 Item Features

One of the most commonly considered item features in existing serendipity-oriented RSs is *item (un)popularity*, given that unpopular items are likely to be unknown by users and hence assumed to be more unexpected than popular ones. For example, [Lu et al. \(2012a\)](#) revised the optimization function of the Matrix Factorization (MF) by adding a popularity term to lower the importance of popular items. [Wang et al. \(2018\)](#) incorporated item popularity to identify active users in the system and further attached a higher weight to active users’ ratings when computing predicted ratings through typical collaborative filtering.

The second type of frequently used item feature is *item (dis)similarity to the user profile*. It was assumed that the more different/dissimilar an item is from the set of items that the user has visited before, the more unexpected it would be ([Zheng et al., 2015](#)). For example, [Zheng et al. \(2015\)](#) combined both item unpopularity and dissimilarity to define unexpectedness and proposed an unexpectedness-augmented PureSVD latent factor model. [Kotkov et al. \(2018b\)](#) also considered both item unpopularity and

dissimilarity in their proposed hybrid re-ranking algorithm. In these two works (Zheng et al., 2015; Kotkov et al., 2018b), popularity was defined based on the number of user visits, and item dissimilarity was computed as the average of pairwise item dissimilarities between the currently considered item and those in the user profile. Nakatsuji et al. (2010) proposed a class distance metric based on item taxonomy to compute item similarity, that is, the smallest distance from the target item’s class to the classes that the user has recently accessed. Adamopoulos and Tuzhilin (2014) proposed a utility model to calculate an item’s unexpectedness, considering the distance of the target item from a set of expected items. More recent serendipity-oriented recommendation approaches, such as those based on graphs (Onuma et al., 2009; De Gemmis et al., 2015; Tuvail, 2019) or neural networks (Pandey et al., 2018; Li et al., 2020), implicitly leveraged the item (dis)similarity in the process of recommendation generation.

The third type of item feature is *temporal information*. For example, Chen et al. (2019b) verified that timeliness (i.e., timely recommendation) has a positive effect on users’ perceived serendipity. Chiu et al. (2011) employed the access time of users to the target item as one dimension of measuring serendipity, assuming that an earlier access time to a new item leads to more serendipity. Kawamae (2010) attempted to predict when a user would purchase an item and reduce the time cost for users to find that item. The authors anticipated this time-saving strategy to have an effect of unexpectedness.

The fourth type is *co-occurrence information*, because it was assumed that the higher co-occurrence probability one item has with other items, the lower unexpectedness it will trigger. There are two kinds of co-occurrence information: One is the probability that the target item’s attributes appear in the user’s previously visited items (Akiyama et al., 2010; Niu and Abbas,

2017), and the other is the probability that different items appear in the same user's profile (Huang et al., 2018)). For instance, Akiyama et al. (2010) defined a TV program's unexpectedness by considering the possibility of its attributes' co-occurrence in all programs. Niu and Abbas (2017) assumed items with attributes less likely to appear together in the target user's profile would generate high-level unexpectedness. Huang et al. (2018) considered *context-independent* co-occurrence based on the assumption that the higher the probability of two entities appearing in the same query, the stronger their relatedness, as well as *context-dependent* co-occurrence through a language model that measures the relatedness between two entities' descriptions.

In addition, some papers considered *domain-specific item features*. For instance, Zhang et al. (2012) considered *listener diversity* (the entropy of the artist's listener distribution) to enhance the serendipity of their hybrid rank-interpolation music recommendations. Shen et al. (2020) took advantage of music features (e.g., artist, genre, language, release year, acoustic features, and lyrics) to build their serendipity-oriented RS.

## 2.2 User Characteristics

According to our literature survey, one of the most commonly considered user characteristics in existing serendipity-oriented RSs is *curiosity* (Chen et al., 2019b), an important premise of users' appetite for novel knowledge and experiences in the field of psychology (Berlyne, 1960; Zhao and Lee, 2016). For example, Menk et al. (2017) predicted the target user's curiosity from their data in social networking, and then generated serendipitous recommendations tailored to the predicted curiosity value. Niu and Abbas (2017) proposed a computational serendipity model with the goal of triggering users' curiosity towards the recommended item at the appropriate

time. [Maccatrozzo et al. \(2017\)](#) developed a serendipity model based on a curiosity theory to trigger user curiosity by matching the level of the recommendation's novelty to the target user's coping potential for new things.

Another type of user characteristic, also inherently related to user personality, is whether the user is an *innovator* or not as this can reflect the person's sensitivity to new items ([Kawamae et al., 2009](#)), on the assumption that items purchased by innovators can surprise their followers. Innovators may have two major properties as stated by [Wang et al. \(2018\)](#): high activity and strong ability to discover new items (e.g., taking less time to find unpopular items), and being unlikely to follow the mainstream. In the work of [Wang et al. \(2018\)](#), the recommendations were retrieved from items interacted by the innovators who were nearest to the current user.

Other considered characteristics include *user elasticity* (measured by the diversity of movie genres in the user's profile), for which an approach for elastic serendipity of movie recommendations was proposed ([Li et al., 2019](#)). Specifically, as a higher degree of user elasticity may indicate the user's higher tendency to accept movies of different genres, movies with lower relevance but higher diversity could be considered as recommendation candidates for such users. The difference between users' long-term and short-term preferences has also been leveraged in recent work. For instance, [Li et al. \(2020\)](#) regarded items not apparently familiar to the user but that meet the user's short-term demands as well as being related to the user's long-term preference to be highly serendipitous.

## 2.3 Limitations of Related Work

Although various types of item features or user characteristics have been considered in existing serendipity-oriented RSs, their effects have been largely based on researchers' assumptions and simply tested by offline simulations. To the best of our knowledge, there has been little work to validate empirically their impacts on users' perceptions of recommendation serendipity. Given that serendipity essentially embodies the pleasantly unexpected feelings of users, the lack of validation with real users may impair the applicability of these systems in real-life situations. Moreover, the majority of related work has ignored the differences likely to be brought by domain specialties (Burke and Ramezani, 2011).

In our work, we primarily analyzed the effects of three types of item features on users' serendipity perception in two application domains: movie recommendations and e-commerce recommendations. The features include item-driven features (e.g., *popularity*), profile-driven features (e.g., *in-profile diversity*), and interaction-driven features (e.g., *attribute co-occurrence*). For user characteristics, we focused on both curiosity and personality (defined by standard psychometric measurements) in our analyses. In addition, we considered some demographic properties (e.g., age and gender) as they were found to have an effect on users' general behaviors and preferences (Hu and Pu, 2013; Schedl et al., 2015).

## 3 User Survey Datasets

We employed two user survey datasets to conduct the analysis: One was collected by Kotkov et al. (2018a) through Movielens (a popular movie recommender system developed by the GroupLens research lab at the

**Table 1** Statistics of MovieLens Serendipity Dataset

<b>User survey data</b>	
Duration of user survey	2017-04-01 to 2018-01-15
# of users in the survey	467
# of (u,i) pairs	2,019
Data type of serendipity feedback	normalized within the range [1,5]
<b>User profile data</b>	
Duration of each user's profile data	3 months
Type of behaviour records	rating
# of users involved	23,334
# of items involved	38,934
# of behaviour records	2,466,278
Data sparsity	99.73%
Average # of users that per item was rated by	63.35 ( <i>std.</i> = 328.55)
Average # of items that per user rated	105.69 ( <i>std.</i> = 216.65)
<b>Item categories</b>	
# of categories	19
Average # of categories per item	1.79 ( <i>min</i> = 0, <i>max</i> = 10)

University of Minnesota), i.e., the *MovieLens Serendipity dataset*<sup>1</sup>; and the other was collected on Mobile Taobao (a popular mobile e-commerce app in China), i.e., the *Taobao Serendipity dataset*<sup>2</sup> (Chen et al., 2019b; Wang et al., 2020).

### 3.1 MovieLens Serendipity Dataset

In the described experiment of Kotkov et al. (2018a), the members of MovieLens were invited to join the survey from April 1, 2017 to January 15, 2018. Movies that were unpopular (with fewer ratings) but liked by the target user (with higher rating above 3.5 out of 5 in the user's historical records) were presented to the user to obtain feedback of how novel or unexpected the movie was. It was thus performed in a retrospective manner as users were asked to recall their feelings when they first received the recommendation. The released survey dataset contains 480 users' responses to the survey questions. We filtered out

<sup>1</sup><https://grouplens.org/datasets/serendipity-2018/>

<sup>2</sup><https://github.com/greenblue96/Taobao-Serendipity-Dataset>

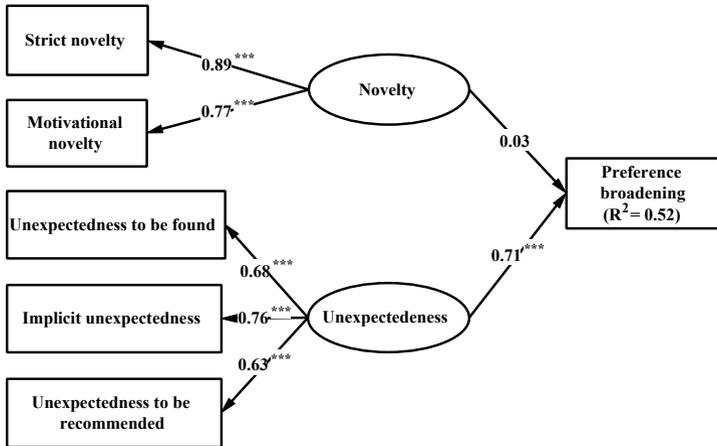
**Table 2** Survey questions related to serendipity in MovieLens Serendipity Dataset (5-point Likert scale for all questions from 1 - “strongly disagree” to 5 - “strongly agree”)

	Mean (Std.)	Median	K-S test
<b>Strict novelty:</b> “The first time I heard of this movie was when MovieLens suggested it to me”	1.710 (1.229)	1	0.367***
<b>Motivational novelty:</b> “MovieLens influenced my decision to watch this movie.”	2.144 (1.334)	2	0.273***
<b>Unexpectedness to be relevant:</b> “I expected to enjoy this movie before watching it for the first time.”	3.683 (0.885)	4	0.283***
<b>Unexpectedness to be found:</b> “This is the type of movie I would not normally discover on my own; I need a recommender system like MovieLens to find movies like this one.”	2.530 (1.262)	2	0.222***
<b>Implicit unexpectedness:</b> “This movie is different (e.g., in style, genre, topic) from the movies I usually watch.”	2.775 (1.176)	3	0.214***
<b>Unexpectedness to be recommended:</b> “I was (or, would have been) surprised that MovieLens picked this movie to recommend to me.”	2.415 (1.073)	2	0.248***
<b>Preference broadening:</b> “Watching this movie broadened my preferences. Now I am interested in a wider selection of movies.”	2.668 (1.104)	3	0.194***

Note: \*\*\* $p < 0.001$  by Kolmogorov-Smirnov test.

cases where less than five movies were rated by a user before they rated the surveyed movie, to avoid a very short length of user profile in our analyses. There are 467 users remaining with totally 2,019 responses (see statistics in Table 1). We then constructed a three-month user profile for each surveyed pair (user, item) that includes the user’s historical rating records in the past three months before s/he rated the surveyed item.

In their survey, participants gave feedback on two novelty questions (*strict novelty* and *motivational novelty*) and four unexpectedness questions (*unexpectedness to be relevant*, *unexpectedness to be found*, *implicit unexpectedness*, and *unexpectedness to be recommended*) (see Table 2). Users were also asked whether the recommendation helped to broaden their preferences (*preference broadening*). *Serendipity* was then defined as various combinations of *novelty* and *unexpectedness*. For instance, one serendipity definition is “*strict serendipity to be found*” that is TRUE if a user’s ratings on both *strict novelty* and *unexpectedness to be found* are both higher than 3 on a 5-point Likert scale. Through cumulative link mixed-effect regression



**Fig. 1** SEM analysis of relationship between variables in Movielens Serendipity Dataset (Note that the value associated with each path is the standardized regression coefficient, \*\*\* $p < 0.001$ , \*\* $p < 0.01$ , and \* $p < 0.05$ ).

analysis, it was found that serendipitous movies according to seven serendipity definitions (except the definition that depends on *motivational novelty* and *unexpectedness to be relevant*) can help broaden user preferences, but no serendipity definition can help increase user satisfaction with the movie (Kotkov et al., 2018a).

However, the survey did not acquire users' perceptions of the recommendation serendipity directly, so it is hard to say which definition more precisely reflects the user's feedback on serendipity itself. Therefore, based on their collected user responses, we aimed to identify a more accurate measurement of user serendipity perception, to accommodate both the *relevance* and *unexpectedness* of the recommended item as discussed before. For this purpose, we re-analyzed their data in the following two steps:

*Step 1:* We performed Principal Component Analysis (PCA). This showed that the correlation coefficients are all smaller than 0.8, suggesting there is no collinearity (Farrar and Glauber, 1967) among the observed variables in Table 2. Moreover, we found that only *unexpectedness to be relevant* exhibited a cross-loading effect (with higher factor loadings on two

latent factors). Considering that it is harmful to user satisfaction as found in the original paper (Kotkov et al., 2018b), we decided to exclude it from the subsequent analysis. As a result, there are two latent factors identified (see Figure 1): “novelty” (related to observed variables *strict novelty* and *motivational novelty*) and “unexpectedness” (related to observed variables *unexpectedness to be found*, *implicit unexpectedness*, and *unexpectedness to be recommended*).

*Step 2:* We then ran the Structural Equation Modeling (SEM) using the SPSS Amos. The final model satisfied several major fit indices:  $\chi^2 = 1.562$  ( $df = 1, p = 0.211$ ),  $CFI = 1.000$ ,  $AGFI = 0.994$ ,  $RMSEA = 0.021$ . There are two major findings according to the final path model (see Figure 1): 1). The impact of “novelty” on *preference broadening* is not significant (coef. = 0.03,  $p > 0.05$ ). 2). “Unexpectedness” has a significantly positive impact on *preference broadening* (coef. = 0.73,  $p < 0.001$ ). Therefore, as the latent factor “novelty” does not engender significant effects, while “unexpectedness” can lead users to broaden their preferences, we finally chose “unexpectedness” as the indicator of users’ unexpected feeling. Moreover, considering that all of the surveyed movies had high relevance to user preference (i.e., the rating was above 3.5 out of 5), it might be reasonable to use the “**unexpectedness**” score to represent their users’ serendipity perceptions. Thus we calculated the “serendipity” score as the average of its three associated observable variables’ values (i.e., *unexp\_find*, *unexp\_imp*, and *unexp\_rec*.)

### 3.2 Taobao Serendipity Dataset

From December 21, 2017 to March 17, 2018, a user survey was conducted on Mobile Taobao (Chen et al., 2019b; Wang et al., 2020). In this survey, we obtained users’ feedback on the recommendation’s serendipity (by asking

**Table 3** Statistics of Taobao Serendipity Dataset

<b>User survey data</b>	
Duration of user survey	2017-12-21 to 2018-03-17
# of users in the survey	11,383
# of (u,i) pairs	11,383
Data type of serendipity feedback	5-point Likert scale
# of users in age groups	18-20: 3,274; 20-30: 4,701; 30-40: 2,433; 40-50: 735; 50-60: 166; >60: 74
# of users in gender groups	female: 7,769; male: 3,614
<b>User profile data</b>	
Duration of each user's profile data	3 months
Type of behaviour records	clicking and purchasing
# of items involved	7,717,420
# of behaviour records	21,405,555 (4.21% purchases)
Data sparsity	99.98%
Average # of users that per item was clicked/purchased by	2.25 ( <i>std.</i> = 6.01) / 0.11 ( <i>std.</i> = 0.72)
Average # of items that per user clicked/purchased	1,802.22 ( <i>std.</i> = 1,557.30) / 72.44 ( <i>std.</i> = 75.36)
<b>Item categories</b>	
# of categories	9,490
Average # of categories per item	3.47 ( <i>min</i> = 2, <i>max</i> = 5)

them to rate the statement directly “*The item recommended to me is a pleasant surprise.*”), and also their demographic properties (including age and gender) and psychological traits (curiosity and Big-Five personality traits; see Table 4). Specifically, user curiosity was assessed via Curiosity and Exploration Inventory-II (CEI-II) (Kashdan et al., 2009), which contains 10 items measuring user “*motivation to seek out knowledge and new experiences*” and “*willingness to embrace the novel, uncertain, and unpredictable nature of everyday life.*” Personality is grounded on the Big-Five factor model and was assessed by the standard Ten-Item Personality Inventory (TIPI) (Gosling et al., 2003) that measures each personality trait via two opposite questions (e.g., “open to new experiences, complex” and “conventional, uncreative” for *Openness to Experience*).

**Table 4** Survey questions in Taobao Serendipity Dataset (serendipity and curiosity questions rated on a 5-point Likert scale from 1 - “strongly disagree” to 5 - “strongly agree”; Big-Five personality questions rated on 7-point Likert scale from 1 - “strongly disagree” to 7 - “strongly agree”)

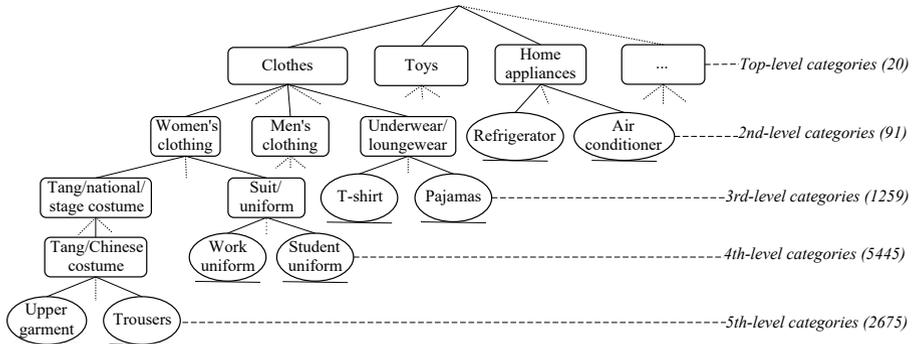
	Mean (Std.)	Median	K-S test
<b>Serendipity:</b> “The item recommended to me is a pleasant surprise.”	2.650 (1.454)	2.00	0.200***
<b>Curiosity:</b> Curiosity and Exploration Inventory-II (CEI-II) (Kashdan et al., 2009)	3.138 (0.819)	3.10	0.035***
<b>Big-Five Personality:</b> Ten-Item Personality Inventory (TIPI) (Gosling et al., 2003)			
-Openness to Experience	4.627 (1.286)	4.50	0.093***
-Conscientiousness	4.562 (1.452)	4.50	0.089***
-Extraversion	4.173 (1.651)	4.00	0.085***
-Agreeableness	4.969 (1.069)	5.00	0.130***
-Neuroticism	4.260 (1.426)	4.00	0.096***

Note: \*\*\* $p < 0.001$  by Kolmogorov-Smirnov test.

We received 13,741 users’ responses. After filtering out invalid answers (e.g., incomplete responses or contradictory ratings to opposite questions) and outliers (e.g., clicked less than five or more than 15,000 items before taking the survey, or related to uncategorized items), 11,383 users’ responses remained, of which 7,769 were from female respondents. We had the past three months’ historical data for each user (i.e., the items they had clicked and purchased before taking part in the survey). Table 3 lists the statistics of user survey data and profile data. Moreover, as shown in Figure 2, the item taxonomy (as provided by *Mobile Taobao*) exhibits a tree structure containing five layers (levels). The leaf category at the end of each path is the direct category that an item belongs to.

Therefore, in comparison with the Movielens Serendipity Dataset, the Taobao Serendipity Dataset not only differs in terms of application domain (e-commerce products vs. movies), but also category structure (hierarchical vs. single-level<sup>3</sup>). Moreover, it contains data of user characteristics.

<sup>3</sup>In the Movielens dataset, there are 19 movie categories (genres) such as “Drama”, “Fantasy”, “Romance”, “Comedy”, etc.



**Fig. 2** The item taxonomy (9,085 leaf categories) of the Taobao Serendipity Dataset.

## 4 Item Features

In this section, we describe how we have classified and extracted item features from the two datasets. The extracted item features can be classified into three groups: *Item-driven*, *profile-driven*, and *interaction-driven*. All the item features are summarized in Table 5.

### 4.1 Item-driven Feature

The item-driven feature considers the information of the target item (i.e., the current recommendation in the survey data)  $r$ , regardless of the target user  $u$ 's profile  $P_u$  (that consists of items that  $u$  has previously interacted with). In other words, it is non-personalized, as the same item's feature values for different users are the same. As mentioned in Section 2, item unpopularity is one of the typical features used to indicate the item's unexpectedness level in related work (Kaminskas and Bridge, 2016), for which the *popularity* is usually determined by the number of users who have interacted with that item.

In the Movielens dataset, we computed the item *popularity* in a similar way, i.e., the percentage of users who have rated the movie. In the Taobao dataset, because the interaction data of all users are not available, we estimated *popularity* through a binary function by using the HOT function provided by

**Table 5** Summary of item features

Feature	Description
<b>Item-driven feature</b>	
<i>Popularity</i>	The item's popularity
<b>Profile-driven feature</b>	
<i>In-profile diversity</i>	The entropy of item categories in the user profile
<b>Interaction-driven feature</b>	
<i>Category-level</i>	
<i>Interaction</i>	The frequency that the user interacted with items of the same category as that of the target item
<i>Day-of-the-week interaction</i>	The frequency that the user interacted with items of the same category on the same day of the week as the target item
<i>Time-of-the-day interaction</i>	The frequency that the user interacted with items of the same category at the same time of the day as the target item
<i>Time difference</i>	The time distance from the target item $r$ to the latest item in the user profile that belongs to the same category
<i>Item-level</i>	
<i>Content-based distance</i>	Jaccard distance based on common content labels of two items
<i>Collaborative-based distance</i>	Jaccard distance based on common users who interacted with the same items
<i>PMI distance</i>	The distance based on Point-wise Mutual Information (PMI) indicating the probability that two items are seen by the same user
<i>Category difference</i>	The category level that two items differ from each other (applicable only to the hierarchical category structure)
<i>Taxonomic distance</i>	The minimum number of hops from one to another item's leaf category (applicable only to the hierarchical category structure)

*Mobile Taobao* (Chen et al., 2019b), which maintains a set of the most popular items. In this way, we measured the target item  $r$ 's popularity as 1 if the item appears in the output of the HOT function, otherwise it is 0.

## 4.2 Profile-driven Feature

Contrary to item-driven features, profile-driven features are dependent of the target user  $u$ 's profile  $P_u$ , but regardless of the target item  $r$ . It should be noted that in the MovieLens dataset, the type of user profile is rating-based

$P_u^{rating}$  (previously rated movies), while in the Taobao dataset, each user has two types of profile: click-based  $P_u^{click}$  (previously clicked items) and purchase-based  $P_u^{purchase}$  (previously purchased items).

We identified one profile-driven feature named ***in-profile diversity***, as inspired by (Wu and Chen, 2015). It particularly reflects the diversity degree of items within the user profile, so the higher the diversity is, the more likely the user is to want to experience different types of items. Formally, we employed entropy (Adomavicius and Kwon, 2012) to define this feature:

$$PD(P_u) = - \sum_{c \in C} \frac{|\{t_{ui} \mid i \in P_u \cap c \in C_i\}|}{|\{t_{ui}\}|} \cdot \log_2 \left( \frac{|\{t_{ui} \mid i \in P_u \cap c \in C_i\}|}{|\{t_{ui}\}|} \right) \quad (1)$$

where  $t_{ui}$  denotes the timestamp when user  $u$  interacted with item  $i$ ,  $|\{t_{ui}\}|$  is the number of the user’s historical records,  $c$  denotes a category from the entire category set  $C$ , and  $C_i$  is the category set of the item  $i$ .  $\frac{|\{t_{ui} \mid i \in P_u \cap c \in C_i\}|}{|\{t_{ui}\}|}$  therefore returns the probability that items belonging to a specific category  $c$  are visited by the target user  $u$ .

In the MovieLens dataset, the category set  $C$  contains 19 categories (e.g., “Drama”, “Fantasy”, and “Romance”). In the Taobao dataset,  $C$  is the set of all leaf categories (9,085 in total) and  $C_i$  is item  $i$ ’s leaf category.

### 4.3 Interaction-driven Features

In addition to features that rely on either the target item  $r$  or the target user  $u$ , the third type of item features depend on the interaction between  $u$  and those items related to  $r$ , called the *interaction-driven features*. According to the granularity of the information used, we divided those features into two groups: *category-level interaction-driven features* and *item-level interaction-driven features*.

### 4.3.1 Category-level Interaction-driven Features

Category-level interaction-driven features inspect the category-level similarity of in-profile ( $P_u$ ) items to the target item  $r$ , for which we define five sub-types according to the different similarity measures considered.

1. **Interaction.** This feature measures the frequency that the user has interacted with items (in their profile  $P_u$ ) of the same category as that of the target item  $r$ :

$$Interaction(r, P_u) = \frac{|\{t_{ui} \mid i \in P_u \cap c_i = c_r\}|}{|\{t_{ui}\}|} \quad (2)$$

In the Movielens dataset,  $c_i = c_r$  means that items  $i$  and  $r$  have at least one category in common. In the Taobao dataset,  $c_i = c_r$  means that  $i$  and  $r$  have the same leaf-level category<sup>4</sup>.

2. **Temporal interaction.** This measures the frequency that the user has interacted with items of the same category as that of the target item at a similar time, given that user preferences can be time-sensitive (Shen et al., 2020; Khoshahval et al., 2018). Taking into consideration the three-month user profile, we derived two time-sensitive features as inspired by (Khoshahval et al., 2018): *day-of-the-week interaction* and *time-of-the-day interaction*.

- **Day-of-the-week interaction:**

$$DoW(r, P_u) = \frac{|\{t_{ui} \mid i \in P_u \cap c_i = c_r \cap day(t_{ui}) = day(t_{ur})\}|}{|\{t_{ui} \mid i \in P_u \cap day(t_{ui}) = day(t_{ur})\}|} \quad (3)$$

where  $day(t_{ui})$  encodes the timestamp  $t_{ui}$  as 0 (Sunday), 1 (Monday), ..., or 6 (Saturday).  $day(t_{ui}) = day(t_{ur})$  means that the day of a week

---

<sup>4</sup>We also computed variations using the top-level categories. Since the results are similar, we have only reported the results of using the leaf-level categories.

22 *Item Features and User Characteristics with Recommendation Serendipity*

when the user interacted with the item  $i$  is the same as the day when the user received the target item  $r$ .

- ***Time-of-the-day interaction:***

$$ToD(r, P_u) = \frac{|\{t_{ui} \mid i \in P_u \cap c_i = c_r \cap period(t_{ui}) = period(t_{ur})\}|}{|\{t_{ui} \mid i \in P_u \cap period(t_{ui}) = period(t_{ur})\}|} \quad (4)$$

where  $period(t_{ui})$  encodes the timestamp  $t_{ui}$  as 1 (morning, from 6 a.m. to 12 p.m.), 2 (afternoon, from 12 p.m. to 6 p.m.), 3 (evening, from 6 p.m. to 12 a.m.), or 4 (night, from 12 a.m. to 6 a.m.);  $period(t_{ui}) = period(t_{ur})$  means that the time of a day when the user interacted with  $i$  is the same as the time when the user received  $r$ .

3. ***Time difference.*** This feature captures the time distance between the target item  $r$  and the latest one in user profile  $P_u$  that belongs to the same category:

$$TimeDiff(r, P_u) = t_{ur} - latest(\{t_{ui} \mid i \in P_u \cap c_i = c_r\}) \quad (5)$$

where  $latest(\cdot)$  returns the most recent timestamp. The result of  $TimeDiff(r, P_u)$  is the numerical time difference in days (e.g., 3.15 days).

### 4.3.2 Item-level Interaction-driven Features

Item-level interaction-driven features are based on pairwise item similarity to reveal the relationship between the target item  $r$  and those in the target user  $u$ 's profile  $P_u$ , for which we also identify five sub-types according to related work ([Kaminskas and Bridge, 2014, 2016](#); [Nakatsuji et al., 2010](#)):

1. **Content-based distance.** This feature is based on the Jaccard distance measure to calculate the similarity of two items based on their content information. We adopted Jaccard distance because it obtained better performances than others (e.g, Cosine distance) in our preliminary experiment (Chen et al., 2019a).

$$ContJacc(r, P_u) = \min_{i \in P_u} 1 - \frac{|C_r \cap C_i|}{|C_r \cup C_i|} \quad (6)$$

where  $C_i$  denotes the set of categories of item  $i$ . This equation therefore returns the minimal content-based Jaccard distance between the target item  $r$  and all items in user profile  $P_u$ . Note that we did not calculate average distance (the same for the following features), because in related work the average distance was found to result in information loss (Kaminskas and Bridge, 2014, 2016)

2. **Collaborative-based distance.**

This feature is based on the ratio of users who accessed both items  $r$  and  $i$  among users who accessed  $r$  or  $i$ :

$$CollabJacc(r, P_u) = \min_{i \in P_u} 1 - \frac{|U_r \cap U_i|}{|U_r \cup U_i|} \quad (7)$$

where  $U_i$  is the set of users who have interacted with item  $i$ .

3. **PMI distance.** Point-wise Mutual Information (PMI) distance aims to check the probability that two items are seen by the same user, under the assumption that if the target item is rarely observed with items in the target user's profile  $P_u$ , it will be more unexpected to the user (Kaminskas and Bridge, 2014, 2016).

$$PMI(r, P_u) = \min_{i \in P_u} \log_2 \frac{p(r, i)}{p(r) * p(i)} / -\log_2 p(r, i) \quad (8)$$

where  $p(i)$  is the probability that item  $i$  is interacted by any user, and  $p(r, i)$  is the probability that two items  $r$  and  $i$  are interacted by the same user.

4. **Category difference.** This feature indicates at which category level the two compared items  $r$  and  $i$  become different over the item taxonomy (applicable only to the hierarchical category structure):

$$CateDiff(r, P_u) = \min_{i \in P_u} (L - |C_r \cap C_i| + 1) \quad (9)$$

where  $L$  is the maximal number of layers (levels) in the item taxonomy (e.g., 5 in the Taobao dataset as shown in Figure 2). Therefore, the largest possible *category difference* is  $L + 1$  (6 in Taobao dataset), when the two items do not share any common category (i.e., are different starting from the top-level category). The smallest is 1, when the two items belong to the same leaf category at the 5-th level.

5. **Taxonomic distance.** This feature is also only applicable to the hierarchical category structure, which is the minimum number of hops from one item  $i$ 's leaf category  $c_i$  to another item  $j$ 's leaf category  $c_j$  over the item taxonomy (called *class distance* in (Nakatsuji et al., 2010)).

$$TaxoDist(r, P_u) = \min_{i \in P_u} hops(c_r, c_i) \quad (10)$$

To illustrate the difference between *category difference* and *taxonomic distance*, suppose there are two items, of which item A belongs to the leaf category “Student uniform” and item B belongs to the leaf category “Work uniform”. The category difference between the two items is  $6 - 3 = 3$ , because they fork at the third level of the Taobao’s item taxonomy

(see Figure 2); while their taxonomic distance is 2 because there are two hops from A’s leaf category “Student uniform” to B’s leaf category “Work uniform”.

## 5 User Characteristics

As mentioned before, we have users’ basic demographic properties and psychological trait values in the Taobao Serendipity Dataset.

### 5.1 Demographic Characteristics

As shown in Table 3, participants of the Taobao user survey were distributed among six *age* groups: 18-20 (3,274 users), 20-30 (4,701 users), 30-40 (2,433 users), 40-50 (735 users), 50-60 (166 users), and above 60 years old (74 users). To simplify the analysis, we divided all users into two age groups: *younger users* who were less than 30 years old (70.06%) and *older users* who were over 30 years old (29.94%).

The other general demographic property is *gender*. There were 7,769 (about 68.25%) female participants and 3,614 (about 31.75%) males. The demographic distributions regarding both age and gender are largely in line with the statistics on online shopping users in China ([Mobile, 2018](#); [Center, 2018](#)).

### 5.2 Psychological Characteristics

Inspired by related work ([Chen et al., 2019b](#); [Maccatrozzo et al., 2017](#); [Zhao and Lee, 2016](#)), we obtained users’ curiosity values in order to analyze their effects on users’ serendipity perceptions. In addition, given that curiosity has been found to significantly correlate with user personality traits, we also

acquired users' personality trait values based on the popularly used Big-Five factor model (Gosling et al., 2003).

1. ***Curiosity***. Curiosity is widely regarded as an important antecedent of users' appetite for new knowledge or experience in the field of psychology (Berlyne, 1960). It can greatly affect the level of pleasure a user may experience when exploring new and unexpected things. We adopted a popularly used curiosity quiz, i.e., the ten-item Curiosity and Exploration Inventory-II (CEI-II) (Kashdan et al., 2009), to measure whether the user has a strong desire for new knowledge or experiences (see description in Section 3). The mean (standard derivation) of Taobao users' curiosity scores is 3.138 (0.819) and the distribution is non-normal according to the Kolmogorov-Smirnov test ( $p < 0.001$ ; see Table 4).
2. ***Big-Five Personality Traits***. The Big-Five factor model (also known as the OCEAN model) is a popularly used taxonomy to define personality (Gosling et al., 2003) from the following five traits:
  - ***Openness to Experience*** indicates a person's facets such as imagination, preference for variety, and intellectual curiosity.
  - ***Conscientiousness*** indicates planned rather than spontaneous behaviors. Conscientious people are dependable and self-disciplined, aiming for achievement.
  - ***Extraversion*** is defined as "an attitude type characterised by concentration of interest on the external object" (Jung, 1983). People with high extraversion tend to be enthusiastic, outgoing, and talkative.
  - ***Agreeableness*** implies facets such as trust, altruism, and tender-mindedness. People scoring high on this trait are empathetic and willing to cooperate.

- **Neuroticism** is concerned with anxiety, hostility, depression, self-consciousness, impulsiveness, and vulnerability (Costa and McCrae, 1992). People who score high on neuroticism are more likely to be moody, feel stressed, and have difficulty in delaying gratification.

The standard Ten-Item Personality Inventory (TIPI) was used to assess participants' Big-Five personality values as obtained in the Taobao dataset (see description in Section 3). Overall, the means (standard derivations) of *Openness*, *Conscientiousness*, *Extraversion*, *Agreeableness*, and *Neuroticism* are 4.627 (1.286), 4.562 (1.452), 4.173 (1.651), 4.969 (1.069), and 4.260 (1.426) respectively, and these traits are not normally distributed among all users ( $p < 0.001$  through the Kolmogorov-Smirnov test; see Table 4).

## 6 Results Analysis

### 6.1 Analysis Methods

To identify the impact of item features on users' perceptions of recommendation serendipity, we employed two non-parametric statistical methods: Spearman's correlation and the Mann-Whitney U test, as users' responses are not normally distributed in both datasets according to the results of the Kolmogorov-Smirnov test (see Tables 2 and 4). Specifically, Spearman's correlation measures the association between two variables based on their rankings, so it is suitable for cases where users' feedback on serendipity was collected using a Likert-scale that is commonly considered as ordinal response (Hildebrand et al., 1977). We also ran the Mann-Whitney U test for each feature between *high* and *low* serendipity groups as divided via a median split method (Iacobucci et al., 2015). In the Movielens dataset, there are 952 (user, item) pairs (47.2%) in the *high* serendipity group (rating

> median 2.33) and 984 (user, item) pairs (48.7%) in the *low* serendipity group<sup>5</sup>. In the Taobao dataset, there are 5,389 users (47.3%; each user rated only one item) in the *high* serendipity group (rating > median 2.0), and 5,994 (52.7%) in the *low* serendipity group. We present the results of the Spearman's correlation and the Mann-Whitney U test in Section 6.2.

For user characteristics that can be defined as categorical variables (e.g., female vs. male, younger vs. older users, highly curious vs. less curious users), we ran the Mann-Whitney U test for the two-group comparison (see results in Section 6.3).

We then fused all the features together into a logistic regression model to investigate the relative importance of each item feature or user characteristic in predicting the user's perceived serendipity. The results are given in Section 6.4.

To further investigate the influence of user profile length, we ran both Spearman's correlation and the Mann-Whitney U test by changing the profile length from 1 week to maximal 12 weeks (i.e., three months). The results are shown in Section 6.5.

## 6.2 Impact of Item Features

From the results in Table 6, we can see that the significance and direction (positive/negative) of Spearman's correlations are basically consistent with the results of the Mann-Whitney U test in each dataset. In the following, we present the results of each type of item features and discuss the similar and different results between the two datasets.

### 1. Item-driven Feature

---

<sup>5</sup>83 (user, item) pairs did not contain answers to all the three unexpectedness questions, so they were excluded from the analysis.

**Table 6** Results of Spearman's correlations (Corr.) and Mann-Whitney U tests (U) with regard to item features

Feature	Movielens		Taobao			
	Rate		Click		Purchase	
	Corr.	U	Corr.	U	Corr.	U
<b>Item-driven feature</b>						
<i>Popularity</i>	<b>-0.118**</b>	L**	<b>-0.138**</b>	L**	<b>-0.138**</b>	L**
<b>Profile-driven feature</b>						
<i>In-profile diversity</i>	<b>0.051*</b>	H*	<b>0.021*</b>	H	<b>-0.029**</b>	L**
<b>Interaction-driven feature</b>						
<i>Category-level</i>						
<i>Interaction</i>	<b>-0.072**</b>	L*	<b>0.057**</b>	H**	<b>0.032**</b>	H**
<i>Day-of-the-week interaction</i>	<b>0.075**</b>	H	<b>0.084**</b>	H**	<b>0.043**</b>	H**
<i>Time-of-the-day interaction</i>	0.029	H	<b>0.080**</b>	H**	<b>0.047**</b>	H**
<i>Time difference</i>	<b>0.048*</b>	H	<b>-0.106**</b>	L**	<b>-0.123**</b>	L**
<i>Item-level</i>						
<i>Content-based distance</i>	0.042	H	<b>-0.079**</b>	L**	<b>-0.022*</b>	L**
<i>Collaborative-based distance</i>	<b>0.104**</b>	H**	<b>-0.113**</b>	L**	-0.002	L
<i>PMI distance</i>	0.002	H	<b>-0.102**</b>	L**	-0.002	H
<i>Category difference</i>			<b>-0.101**</b>	L**	<b>-0.034**</b>	L**
<i>Taxonomic distance</i>			<b>-0.076**</b>	L**	-0.018	L

Note: 1). \*\* $p < 0.01$  and \* $p < 0.05$  for both Spearman's correlation and the Mann-Whitney U test. 2). For the Mann-Whitney U test, the group with higher value is marked (L for the *low* serendipity group and H for the *high* serendipity group).

- **Popularity.** In both the Movielens and Taobao datasets, *popularity* is significantly *negatively* correlated with users' perceived serendipity (e.g.,  $corr. = -0.118$ ,  $p < 0.01$  in the Movielens dataset and  $corr. = -0.138$ ,  $p < 0.01$  in the Taobao dataset). The Mann-Whitney U test reveals that the two groups, *high* and *low* serendipity groups, are significantly different in terms of the item *popularity* (e.g.,  $mean = 0.0124$  vs.  $0.0191$ ,  $p = 0.000$  in the Movielens dataset; and  $0.1895$  vs.  $0.3041$ ,  $p = 0.000$  in the Taobao dataset). The results hence suggest that *less popular items are more likely to be perceived as serendipitous by users.*

## 2. Profile-driven Feature

- **In-profile diversity.** In both the Movielens dataset and the Taobao click-based dataset, the results of Spearman's correlation shows that *in-profile diversity* has a significantly positive correlation

with users' perceived serendipity ( $corr. = 0.051, p < 0.01$  in the Movielens dataset, and  $corr. = 0.021, p < 0.05$  in the Taobao click-based dataset). However, in the Taobao purchase-based dataset, *in-profile diversity* is negatively correlated with users' perceived serendipity ( $corr. = -0.029, p < 0.01$ ). This difference may be explained by the fact that movies are essentially for entertainment consumption (i.e., dominantly hedonic and aesthetic) (Oliver and Raney, 2011; Cooper-Martin, 1991), while online shopping is mainly for practical consumption (i.e., to fulfill utilitarian functions) (Cooper-Martin, 1991; Ford et al., 1988) (see Section 7 for more discussion).

Hence, the interpretation is that *for movie recommendations, the more diverse the movies watched before, the more likely that the user would perceive a new recommendation to be serendipitous. However, for e-commerce recommendations, the more diverse the purchases made, the more difficult it is for the user to feel a new recommendation to be serendipitous.*

### 3. Interaction-driven Features

#### *Category-level Interaction-driven Features:*

- **Interaction.** This feature reflects a user's preference for a certain category. In the Movielens dataset, results of both the Spearman's correlation and the Mann-Whitney U test indicate a significantly negative relationship with users' perceptions of serendipity (e.g.,  $corr = -0.072, p < 0.01$ ), inferring that movie categories that users have interacted less with are more likely to be serendipitous to them. However, the direction is the opposite in the Taobao dataset.

The correlations between *interaction* and users' perceived serendipity in both the click-based dataset and the purchase-based dataset are positive ( $corr. = 0.057$  and  $0.032$  respectively,  $p < 0.01$ ), inferring that in an e-commerce recommender system, users are more likely to be pleasantly unexpected when they receive recommendations of categories that they have frequently interacted with. Thus, *the effects of category-level interaction on users' serendipity perceptions can be different in two product domains.*

- **Temporal interaction.** The results are similar between the two datasets, indicating that recommending an item from a category that users have frequently accessed at the same time point in the past (the day of the week or the time of the day) are more likely to increase users' perceptions of serendipity. Specifically, in the Movielens dataset, Spearman's correlation is significant regarding *Day-of-the-week interaction* ( $corr. = 0.075$ ,  $p < 0.01$ ). In the Taobao dataset, the correlations between users' serendipity perceptions and *Day-of-the-week interaction/Time-of-the-day interaction* are all significant ( $corr. = 0.084$  and  $0.043$  with regard to *Day-of-the-week interaction* in the click-based dataset and the purchase-based dataset respectively; and  $corr. = 0.080$  and  $0.047$  with regard to *Time-of-the-day interaction*). This suggests that *recommendation of items in a timely way can lead users to consider it to be more serendipitous.*
- **Time difference.** In the Movielens dataset, time difference is positively correlated with users' perceived serendipity ( $corr. = 0.048$ ,  $p < 0.05$ ), indicating that the more recently the

current recommendation's category has been visited by the target user, the less likely the user is to consider it serendipitous. However, in the Taobao dataset, the results of both the Spearman's correlation and the Mann-Whitney U test show that if the category of the currently recommendation has been recently clicked or purchased, it may be more serendipitous to the user. In the click-based Taobao dataset the mean of *time difference* is 9.24 days in the high serendipity group vs. 12.12 days in the low group ( $corr. = -0.106, p < 0.01$ ); and in the purchase-based Taobao dataset it is 76.18 days vs. 79.54 days ( $corr. = -0.123, p < 0.01$ ). The finding in the Taobao dataset contradicts our common assumption that an unexpected item would be from a category that the user has not recently accessed (Kotkov et al., 2016). Thus, this infers that *the effects of time difference on users' serendipity perceptions can be different in two product domains.*

**A short summary.** Through the above analyses, we find: 1). the effects of two temporal interaction features (i.e., *day-of-the-week interaction* and *time-of-the-day interaction*) are similar in the two datasets, i.e., timely recommendations in line with the user's preferences are likely to be perceived as serendipitous. 2). *Interaction* and *time difference* affect users' serendipity perceptions in different ways for movie recommendations and e-commerce recommendations. Specifically, users are more likely to perceive serendipity in movie categories that they have rarely watched (i.e., low *interaction*) or have not watched for a long time (i.e., large *time difference*); whereas, when shopping online,

users will be likely to perceive serendipity in products from categories they have frequently or recently visited.

***Item-level Interaction-driven Features:***

- **Content-based distance.** There is no significant relationship between this feature and users' serendipity perceptions in the Movielens dataset, but in the Taobao dataset both the results of the Spearman's correlation and the Mann-Whitney U test show that items with shorter *content-based distance* are more likely to be perceived as highly serendipitous ( $corr. = -0.079$  and  $-0.022$  respectively in click-based and purchase-based datasets,  $p < 0.01$ ). Thus, *recommending an e-commerce item with a shorter content-based distance to the target user's profile can help increase its perceived serendipity.*
- **Collaborative-based distance.** *Collaborative-based distance* acts differently between the two datasets. It shows that the larger the distance, the more serendipitous the item is to users in the Movielens dataset ( $corr. = 0.104$ ,  $p < 0.01$ ), but the opposite is the case in the Taobao dataset ( $corr. = -0.113$ ,  $p < 0.01$  for the click-based dataset). Thus, *the effects of collaborative-based distance on users' serendipity perceptions can be different in two product domains.*
- **PMI distance.** This feature takes effects only in the click-based Taobao dataset. The results show that items with a smaller PMI distance to the user profile are more likely to be perceived of high serendipity, verifying the previous assumption (Kaminskas and Bridge, 2014, 2016) *in the e-commerce domain that a*

*recommendation that has rarely been observed with items in the user's profile will be more serendipitous to the user.*

- **Category difference.** This feature, and also the next one, *taxonomic distance*, is only applicable to the Taobao dataset for which a hierarchical item taxonomy is available (see Figure 2). It shows that the minimal category differences (regarding both click-based and purchase-based profiles) in the *high* serendipity group are significantly shorter than those in the *low* group (e.g.,  $mean = 0.4118$  vs.  $mean = 0.4445$  in the click-based dataset,  $p < 0.01$ ), implying that the serendipitous recommendation is not necessarily of a bigger category difference from those the user has interacted with. Thus, *the category difference of a highly serendipitous item from the user profile can be shorter.*
- **Taxonomic distance.** This refers to the number of hops from one item's leaf category to that of another item. It is interesting to find that the correlation between users' serendipity perceptions and *taxonomic distance* in the click-based dataset is significantly negative ( $corr. = -0.076$ ,  $p < 0.01$ ), indicating that *the taxonomic distance from the highly serendipitous item to the click-based user profile is also shorter (i.e., traversing fewer nodes).*

**A short summary.** We find that the movies that users are not familiar with (based on *collaborative-based distance*) will bring more serendipity, while for e-commerce, users tend to perceive recommendations more relevant to their preferences (i.e., with shorter *content-based distance*, shorter *collaborative-based distance*, shorter *PMI*

**Table 7** Results of the Mann-Whitney U test with regard to psychological characteristics

Characteristic	# users in the low (high) group	Mann-Whitney U	Mean of serendipity perception in the low (high) group
<b>Curiosity</b>	5,828 (5,555)	20,213,781.000**	2.33 ( <b>2.99</b> )
<b>Big-Five personality</b>			
<i>Openness to Experience</i>	6,187 (5,196)	16,742,037.000**	2.60 ( <b>2.72</b> )
<i>Conscientiousness</i>	6,280 (5,103)	17,908,544.000**	2.51 ( <b>2.83</b> )
<i>Extraversion</i>	6,190 (5,193)	16,863,659.500**	2.59 ( <b>2.72</b> )
<i>Agreeableness</i>	5,246 (6,137)	15,293,260.000**	<b>2.72</b> (2.59)
<i>Neuroticism</i>	5,956 (5,427)	17,437,065.000**	2.55 ( <b>2.76</b> )

Note: \*\* $p < 0.01$  and \* $p < 0.05$  for the Mann-Whitney U test.

*distance*, smaller *category difference*, and shorter *taxonomic distance*) as being of greater serendipity.

### 6.3 Impact of User Characteristics

This part of the analysis was only conducted in the Taobao dataset that contains information of user characteristics.

#### 1. Demographic Characteristics

- **Age.** Because both age groups (younger and older) and gender (female and male) were not balanced (see Section 3.1), we conducted the Mann-Whitney U test to analyze their effects as it makes no assumption about the data distribution ([Wikipedia Contributors, 2022](#)). The results show that the serendipity level perceived by younger users is significantly lower than that of older users, implying that *older people are more inclined to feel a recommendation to be serendipitous*.
- **Gender.** For *gender*, we find that male users gave significantly higher serendipity scores on recommendations than female users did (2.79 vs. 2.59,  $p < 0.001$ ). This infers that *men are more inclined to feel a recommendation to be serendipitous*.

#### 2. Psychological Characteristics

- **Curiosity.** We first applied the median split method (Iacobucci et al., 2015) to divide all users into two curiosity groups (i.e., *high* and *low*). The results show that people with *higher curiosity* is likely to perceive a recommendation to be more serendipitous (see Table 7), validating the previous findings that *highly curious users are more likely to enjoy novel and unexpected items* (Kotkov et al., 2016; Chen et al., 2019b).
- **Big-Five personality traits.** In respect of each Big-Five personality trait, we split all users into two groups, e.g., *high* Openness and *low* Openness, still through the median split method. It shows people of *high* Openness to Experience, Conscientiousness, Extraversion, or Neuroticism gave significantly higher scores on item serendipity, while those of *high* Agreeableness gave significantly lower scores.

We further calculated the correlations between user curiosity and the Big-Five personality traits. Openness has the largest positive correlation with curiosity ( $corr. = 0.410$ ), followed by Extraversion (0.276), Conscientiousness (0.157), and Neuroticism (0.117), while Agreeableness has a negative correlation with curiosity (-0.175). This infers that *highly curious users, who by nature are more open, more conscientious, more extrovert, and more neurotic, are more inclined to perceive a recommendation as highly serendipitous.*

## 6.4 Relative Importance of Features

So far we have analyzed the effects of individual item features and user characteristics. It should be interesting to identify their relative importance in influencing users' serendipity perceptions when being considered together.

**Table 8** The relative importance of considered features (number in bracket indicates the importance ranking position)

	Feature	Movielens	Taobao	
			Click	Purchase
Item Feature	<i>Popularity</i>	0.799*** (1)	0.741*** (3)	0.776*** (5)
	<i>Day-of-the-week interaction</i>	1.127** (2)		
	<i>Category difference</i>		0.826*** (4)	0.737*** (4)
	<i>Interaction</i>		0.905*** (7)	0.944** (12)
	<i>Time difference</i>		0.912*** (8)	0.837*** (8)
	<i>Time-of-the-day interaction</i>		1.068** (10)	
	<i>Collaborative-based distance</i>		0.948* (12)	
	<i>Content-based distance</i>			1.286* (3)
	<i>Taxonomic distance</i>			1.198* (6)
	<i>In-profile diversity</i>			0.911*** (11)
User Characteristic	<i>Curiosity</i>		<b>1.769***</b> (1)	<b>1.717***</b> (1)
	<i>Age</i>		1.361*** (2)	1.380*** (2)
	<i>Gender</i>		0.855** (5)	0.833*** (7)
	<i>Openness</i>		0.897*** (6)	0.895*** (9)
	<i>Conscientiousness</i>		1.088*** (9)	1.090*** (10)
	<i>Neuroticism</i>		1.060* (11)	1.051** (13)

Note: Only features with significant coefficients are listed in the table. The reported value is the odds ratio  $EXP(B)$  (Pantouvakis, 2010) with significance: \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ .

To this end, we built an ordinal logistic regression model that involves all the item features (and user characteristics for the Taobao dataset) as independent variables and users' serendipity perceptions as the dependent variable:

$$serendipity \sim F_1 + F_2 + \dots + F_{|F|} + C_1 + C_2 + \dots + C_{|C|} \quad (11)$$

where  $F$  denotes the set of item features, and  $C$  denotes the set of user characteristics if any.

*popularity* and *day-of-the-week interaction* are two significant predictors (among all considered item features) of serendipity. Between them, the importance of *popularity* is relatively higher (odds ratio = 0.799 vs. 1.127<sup>6</sup>).

In the click-based Taobao dataset, there are six item features and six user characteristics (out of a total of 19 features) that are significant predictors of users' serendipity perceptions. Among them, *curiosity* has the largest importance, followed by *age*, *popularity*, *category difference*, and *gender*. In the purchase-based Taobao dataset, seven item features and six user characteristics are significant predictors, among which *curiosity* is still the most important, followed by *age*, *content-based distance*, *category difference*, and *popularity*.

There are two major observations regarding the Taobao dataset: 1). Considering the results reported in the previous section (see Table 6), it can be seen that although some item features (e.g., *interaction* and *collaborative-based distance*) have significant relationships with serendipity, the effects can be weak compared to other features (e.g., *popularity* and *category difference*) in the relative importance analysis. 2). Compared with *interaction-driven* item features, user characteristics (especially *curiosity* and *age*) and item-driven features (i.e., *popularity*) appear to be more important predictors of users' serendipity perceptions.

## 6.5 Influence of Profile Length

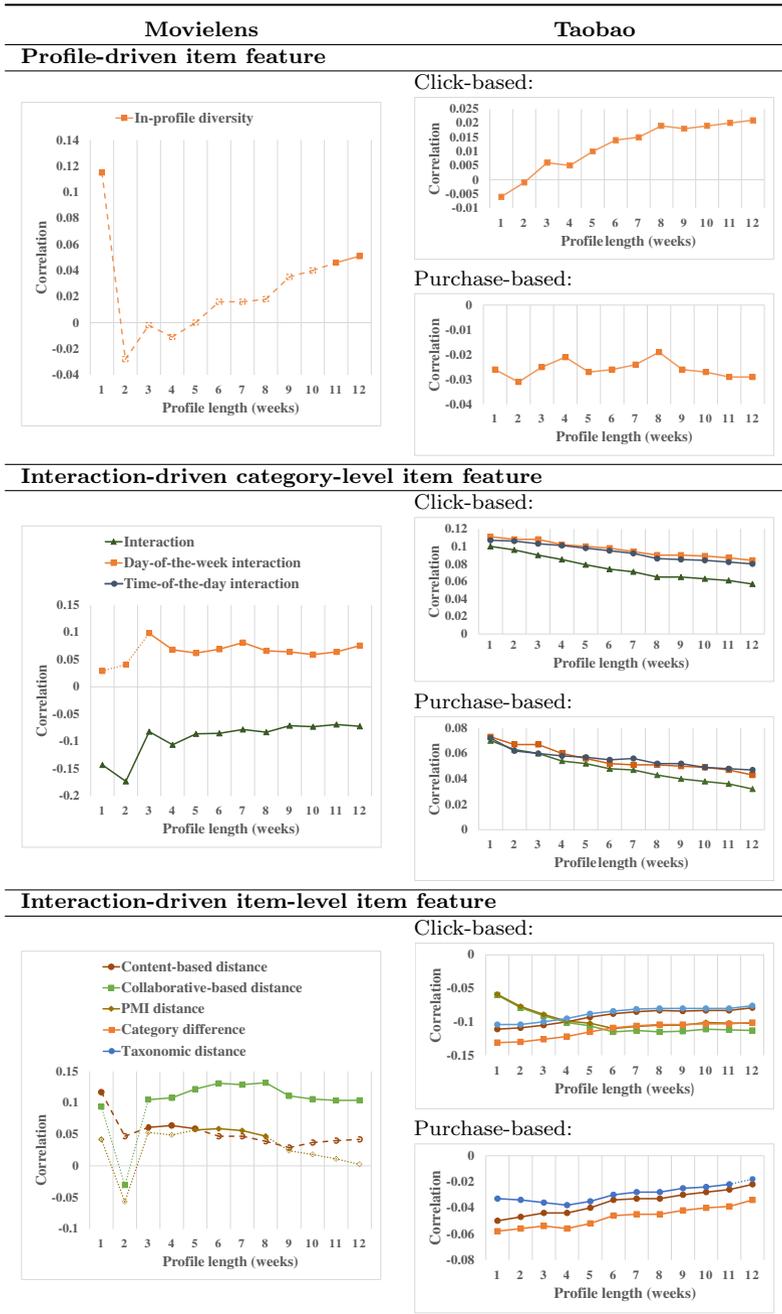
Another question we were interested in addressing is how the results vary in different lengths of user profile, given that *profile-driven* and *interaction-driven* item features can both be affected by the size of user profile (e.g., a 1-week vs. 12-week profile). For this purpose, we calculated Spearman's correlations for 12

---

<sup>6</sup>The values indicate that, if *popularity* rises by one unit, users' perceived serendipity will decrease by  $|1 - 0.799| * 100\% = 20.1\%$ , while one-unit increase of *day-of-the-week interaction* will lead to  $|1 - 1.127| * 100\% = 12.7\%$  increase of users' perceived serendipity.

different lengths ranging from one to 12 weeks (maximally three months in our datasets). The results are shown in Table 9. Note that: 1). *Time difference* was not considered because it only considers the last time that the user accessed the same category as that of the recommendation, and 2). only features that have significant results in at least one profile length setting are illustrated in the figures.

It can be seen that the correlations between *interaction-driven* item features and users' serendipity perceptions are stronger with a short-term user profile (e.g., a 3-week profile) than those with a long-term profile (e.g., a 12-week profile). For example, the correlation coefficients of category-level interaction-driven features (such as *interaction* and *time-of-the-day interaction* in the Movielens dataset, and *interaction*, *day-of-the-week interaction*, and *time-of-the-day interaction* in both the Taobao click-based and purchase-based datasets) with users' perceived serendipity decrease with the increase of profile length (e.g., from  $corr. = 0.100$  using the 1-week user profile to  $corr. = 0.057$  with the 12-week profile, regarding *interaction* in the Taobao click-based dataset). Moreover, in the Movielens dataset, the performances of *category-level interaction-driven* features are more stable than *item-level interaction-driven* features. For example, the feature *interaction* keeps significant correlations with users' perceptions of serendipity, no matter which profile length is considered. However, as the length becomes longer, the correlation between the feature *PMI distance* and user serendipity perception first changes from non-significant to significant (in the 5-week profile), and then back to non-significant (in the 9-week profile). The latter fluctuation is unfavorable in practical applications, because when the data are changed, the feature's validity may not be guaranteed. In the Taobao dataset, all the *interaction-driven* features (both

**Table 9** The influence of user profile length

*Note:* 1). Only features that have significant correlations with users' perceptions of serendipity in at least one profile length setting are displayed in the diagram. 2). The dashed line indicates there are not significant correlations in that part.

category-level and item-level) are significantly correlated with users' perceived serendipity regardless of the profile length.

However, with the *profile-driven* feature, i.e., *in-profile diversity*, the longer the profile length is, the richer the information there can be, so the correlation becomes stronger. To be more specific, in the Movielens dataset, *in-profile diversity* is significantly correlated with user serendipity perception when using the 11-week or 12-week profile ( $corr. = 0.046$  and  $0.051$  respectively,  $p < 0.05$ ), but there is no significant relationship when the profile length is shorter than 11 weeks. In the Taobao dataset, especially when using the click-based profile, the advantage of using long-term profiles is more obvious. The correlation coefficient grows as the profile becomes longer, and reaches the maximum value of 0.021 when using the 12-week data.

**A short summary.** The results show that the shorter the profile is, more strongly *interaction-driven* features correlate with users' perceived serendipity, but for the *profile-driven* item feature, the longer the profile is, more strongly it correlates with users' perceived serendipity.

## 7 Discussion

### 7.1 Major Findings

In this section, we summarize our major findings in response to the research questions raised at the beginning.

This study investigated the effects of a number of item features on user perceptions of recommendation serendipity with two user survey datasets: the Movielens Serendipity Dataset (Kotkov et al., 2018a) containing 467 users' serendipity feedback and the Taobao Serendipity Dataset (Chen et al., 2019b; Wang et al., 2020) with 11,383 users' feedback. The analyses disclose *domain-independent* and *domain-specific* observations (see the summary of

**Table 10** Summary of major findings with regard to item features and user characteristics

<b>Item feature</b>	
<b>Item-driven feature (Movielens and Taobao datasets)</b>	
<i>Popularity</i>	Less popular items are more likely to be perceived as serendipitous.
<b>Profile-driven feature (Movielens and Taobao datasets)</b>	
<i>In-profile diversity</i>	For movie recommendations, the more diverse the movies watched before, the more likely that the user would perceive a new recommendation to be serendipitous. However, for e-commerce recommendations, the more diverse the purchases made, the more difficult it is for the user to feel a new recommendation to be serendipitous.
<b>Interaction-driven feature</b>	
<b>Category-level (Movielens and Taobao datasets)</b>	
<i>Interaction</i>	For movie recommendations, movie categories that users have interacted less with are more likely to be serendipitous to them. However, the opposite is the case for e-commerce recommendations.
<i>Temporal interaction</i>	Recommendation of items in a timely way can lead users to consider it to be more serendipitous.
<i>Time difference</i>	For movie recommendations, the more recently the current recommendation's category has been visited by the target user, the less likely the user is to consider it serendipitous. However, the opposite is the case for e-commerce recommendations.
<b>Item-level (Movielens and Taobao datasets, except those with # applicable only to Taobao dataset)</b>	
<i>Content-based distance</i>	Recommending an e-commerce item with a shorter content-based distance to the target user's profile can help increase its perceived serendipity.
<i>Collaborative-based distance</i>	For movie recommendations, the larger the distance is, the more serendipitous the item is to users. However, the opposite is the case for e-commerce recommendations.
<i>PMI distance</i>	An e-commerce recommendation that has rarely been observed with items in the user's profile will be more serendipitous to the user.
<i>Category difference#</i>	The category difference of a highly serendipitous item from the user profile is shorter.
<i>Taxonomic distance#</i>	The taxonomic distance from the highly serendipitous item to the click-based user profile is shorter.
<b>User characteristic (applicable only to Taobao dataset)</b>	
<b>Demographic characteristic</b>	
<i>Age</i>	Older people are more inclined to feel a recommendation to be serendipitous.
<i>Gender</i>	Men are more inclined to feel a recommendation to be serendipitous.
<b>Psychological characteristic</b>	
<i>Curiosity</i>	Highly curious users are more likely to enjoy serendipitous items.
<i>Big-Five personality traits</i>	People who are of high Openness to Experience, Conscientiousness, Extraversion, or Neuroticism, or low Agreeableness are more inclined to perceive a recommendation as highly serendipitous.

all major observations in Table 10). Specifically, we find that some features have consistent results across the two datasets: 1). Lower *item popularity* is significantly correlated with higher recommendation serendipity as perceived by the user. 2). The results about *day-of-the-week interaction* and *time-of-the-day interaction* show that recommendations that meet users' daily or weekly behavior patterns may lead to a sense of serendipity.

Several features, however, have different effects in the two datasets: 1). Users are more likely to experience serendipity from movies in categories that they have rarely watched (i.e., lower *interaction*) or have not watched for a long time (i.e., larger *time difference*), or movies that differ from previous views (i.e., longer *collaborative-based distance*). 2). In the e-commerce domain, users experience more serendipity from products in categories that they have frequently (i.e., higher *interaction*) or recently (i.e., smaller *time difference*) clicked or purchased, or from products similar to their previous clicked or purchased items (i.e., smaller *content-based distance*, *collaborative-based distance*, *PMI distance*, *category difference*, and *taxonomic distance*).

These differences imply that users may place different emphases on the two components of serendipity, i.e., *relevance* and *unexpectedness*, in different product domains. To be specific, it seems that users focus more on the *unexpectedness* component when choosing movies, and more on the *relevance* component when searching for e-commerce products. According to Maslow's hierarchy of needs (Maslow, 1943), humans' needs can be classified into five levels from physiological needs (e.g., food and water), safety needs (e.g., security), belongingness and love needs (e.g., family and friends), esteem needs (e.g., feelings of accomplishment), to self-actualization (e.g., conducting creative activities). In the case of movie recommendations (part of fulfilling users' self-actualization needs), users might be inclined to explore more diverse items for obtaining new experiences, while for e-commerce products that mainly serve their daily physiological and safety needs, users might have lower tolerance to less relevant recommendations.

Moreover, with the Taobao Serendipity Dataset, we compared the results based on click-based user profiles and purchase-based profiles. The results

indicate that, other than the *profile-driven* feature (i.e., *in-profile diversity*), other features that are significantly correlated with users' serendipity perceptions behave similarly in the two types of profile datasets. For example, the correlations of *day-of-the-week interaction* with users' perceived serendipity in both datasets are positive (0.084 and 0.043 respectively,  $p < 0.01$ ). This also indicates that regarding *in-profile diversity* users who purchased less diverse categories tend to perceive higher serendipity. However, this is less obvious in the click-based dataset. It hence suggests that diversity might be better measured by using their purchase records given that clicking behavior could be aimless when shopping ([Wikipedia Contributors, 2021](#)).

In the Taobao dataset, we were also able to analyze the relationships between several users' characteristics and their perceived sense of serendipity (see [Table 10](#)). Both *age* and *gender* induced significant effects: older (above 30 years old) or male users are more likely to feel a recommendation to be serendipitous, while younger or female users are harder to feel unexpected. In addition, our results verify that users with higher *Curiosity/Openness/Conscientiousness/Extraversion/Neuroticism*, or lower *Agreeableness*, are more likely to give a higher score on the recommendation's serendipity. The reason for this may be that the five personality traits are all related to curiosity to a certain extent ([Kashdan et al., 2009](#)).

We further analyzed the relative importance of all considered features (including user characteristics) showing that 1). In both Movielens and Taobao datasets, *item popularity* is relatively of high importance, and 2). in the Taobao dataset containing both item features and user characteristics, user characteristics especially *Curiosity* and *Age* have greater importance than item features. Moreover, no matter which type of user profile

(click-based or purchase-based) was considered, five user characteristics (i.e., *Age*, *Gender*, *Curiosity*, *Openness*, *Conscientiousness*, and *Neuroticism*) and four item features (i.e., *item popularity*, *category difference*, *interaction*, and *time difference*) are all significant predictors of users' perceived serendipity.

In addition, the analysis of the effect of profile length (varying from 1 to 12 weeks) reveals that in both the Movielens and Taobao datasets, *interaction-driven* features (especially *interaction*, *day-of-the-week interaction*, *content-based distance*, *collaborative-based distance*, and *PMI distance*) have stronger correlations with users' perceived serendipity when using a shorter-term profile. This could be because factors leading to serendipity may change from time to time and therefore recent interactions are more important to the calculation of *interaction-driven* features. However, two features were differently affected by the profile length in the Movielens and Taobao datasets. One is *time-of-the-day interaction*, which, when being correlated with users' serendipity perceptions with a certain length of user profile in the Taobao dataset, had no significant correlation in the Movielens dataset. The other is *in-profile diversity*, which was significantly correlated with users' serendipity perceptions in the Movielens dataset only when a longer-term profile was used (i.e., 11-week or 12-week profile); whereas in the Taobao dataset, it showed significant correlations regardless of the profile length.

## 7.2 Practical Implications

The findings of this study could be constructive in the design of more effective serendipity-oriented recommendation algorithms. For instance, *item popularity* is a domain-independent feature that negatively correlates with users' perceived serendipity in both the Movielens and Taobao datasets. This

suggests that the methods with the objective of achieving recommendation serendipity based on this feature (Lu et al., 2012a; Kawamae, 2010; Zheng et al., 2015) could be generalized to different domains (at least movies and e-commerce products). Moreover, some special features considered in this work, such as *day-of-the-week interaction* that is also domain-independent in terms of influencing users' serendipity perceptions, could be usefully incorporated into a classical recommendation framework to strengthen its serendipity. For example, the method based on SVD++ to generate serendipitous recommendations (Lu et al., 2012b; Zheng et al., 2015) could further be enhanced by adding a vector representing the user's day-of-the-week preference for each latent factor.

On the other hand, care should be taken in utilizing *domain-specific* features. For instance, although in both datasets we considered a fixed 3-month user profile to extract those item features, users' activity degrees actually varied between the two datasets. Taobao users were more active than Movielens users according to the average number of visited items. This may partially explain why some features, such as *time-of-the-day interaction* measuring the frequency that the user has interacted with items on the same day as the target item, give significant correlation with users' serendipity perceptions in the Taobao dataset, but not in the Movielens dataset. However, other features such as *time difference* that do not rely on the activity degree might be directly used to build domain-specific recommendation algorithms for reflecting users' serendipity preferences.

Our findings also verified the impacts of several user characteristics, especially *curiosity* and *big-five personality traits*, on users' perceptions of recommendation serendipity. This implies that related methods (Shen et al., 2020; Kotkov et al., 2016) based on such characteristics might achieve the

serendipity objective from the users' perspective. However, because our validation was done with only one dataset due to the lack of user data in other domains, more studies might be needed when employing those methods in real situations.

### 7.3 Limitations of Our Work

There are two major limitations of this work. *Firstly*, although the study investigated 11 item features and 8 user characteristics, the findings suggest that the interaction between some features, such as *time difference* and *category difference*, are worth being investigated in-depth, because when they are considered together, the combined effect may be stronger than that of treating them separately. Besides, because of the limited data size, we only measured the influence of profile length up to 12 weeks, but as the results of *time-of-the-day interaction* imply, its correlation with users' serendipity perceptions might further increase with the increased profile length.

*Secondly*, the two datasets, although collected in two different product domains, differ in terms of other potential factors such as users' personal background, suggesting that those identified domain differences might not be simply explained by the domain specialty itself. Ideally, the comparison should be done with the same group of users, but as this is not feasible due to the unavailability of such datasets, we expect that more studies could be done in the future to address this issue.

## 8 Conclusions

In order to understand how item features and user characteristics may affect users' perceptions of recommendation serendipity, in this study we analyzed two publicly accessible user survey datasets: the Movielens Serendipity Dataset

(Kotkov et al., 2018a) and the Taobao Serendipity Dataset (Chen et al., 2019b; Wang et al., 2020). Through various statistical analyses, we have not only identified the significant effects of several types of features (i.e., *item-driven*, *profile-driven*, and *interaction-driven*), but also analyzed all of the features' relative importance, as well as the effects of user characteristics in the Taobao dataset and the influence of user profile length in both datasets.

Concretely, in response to the three research questions we raised at the beginning, there are several interesting findings: 1). Regarding what item features can be correlated with users' perceived serendipity, we found that two particular item features (i.e., *popularity* and *temporal interaction*) perform similarly between the Movielens and Taobao datasets. However, three features' effects are opposite in the two domains (i.e., *in-profile diversity*, *interaction*, and *time difference*), and three only take significant effects in one dataset (i.e., *content-based distance* and *PMI distance* for e-commerce items, and *collaborative-based distance* for movie items). We further investigated the differences between the two types of Taobao dataset, i.e., click-based and purchase-based, in terms of those features' exact roles. 2). As for what user characteristics can affect their perceived serendipity that was only analyzable in the Taobao dataset, we found that older people, men, or more curious users (who are essentially more open, more conscientious, more extrovert, and more neurotic in terms of their Big-Five personality traits) are more inclined to perceive a recommendation as highly serendipitous. 3). When putting all of the considered features together to predict users' serendipity perception, it shows that, in the Movielens dataset, *popularity* and *day-of-the-week interaction* are two significant predictors; while *curiosity*, *age*, *popularity*, and *category difference* are more predictive in the Taobao dataset. 4). Moreover, the features' correlation with users'

serendipity perception can vary in different lengths of user profile. In particular, it shows that the shorter the profile is, more strongly interaction-driven features (e.g., *interaction* and *time-of-the-day interaction*) correlate with users' perceived serendipity, but for longer user profiles, the profile-driven item feature (i.e., *in-profile diversity*) is more strongly correlated with users' perceived serendipity.

We discussed the reasons behind those findings and their practical implications to develop more effective serendipity-oriented recommendation algorithms in consideration of the dataset's domain properties. In the future, we will be engaged in addressing the limitations noted in Section 7.3. We will also be interested in incorporating those effective features into the implementation of a real serendipity-oriented recommender system and measuring the system's practical performance in comparison with the state-of-the-art methods.

## References

- Adamopoulos P, Tuzhilin A (2014) On Unexpectedness in Recommender Systems: Or How to Better Expect the Unexpected. *ACM Transactions on Intelligent Systems and Technology (TIST)* 5(4):54:1–54:32, DOI 10.1145/2559952, URL <http://doi.acm.org/10.1145/2559952>
- Adomavicius G, Kwon Y (2012) Improving Aggregate Recommendation Diversity Using Ranking-Based Techniques. *IEEE Transactions on Knowledge and Data Engineering* 24(5):896–911, DOI 10.1109/TKDE.2011.15
- Akiyama T, Obara K, Tanizaki M (2010) Proposal and Evaluation of Serendipitous Recommendation Method Using General Unexpectedness. In:

Workshop on the Practical Use of Recommender Systems, Algorithms and Technologies (PRSAT 2010), held in conjunction with RecSys 2010

Berlyne DE (1960) *Conflict, Arousal and Curiosity*. McGraw-Hill

Bogers T, Björneborn L (2013) Micro-serendipity: Meaningful Coincidences in Everyday Life Shared on Twitter. In: *Proceedings of the iConference 2013*, pp 196–208, DOI 10.9776/13175

Burke R, Ramezani M (2011) Matching Recommendation Technologies and Domains, Springer US, Boston, MA, pp 367–386. DOI 10.1007/978-0-387-85820-3\_11, URL [https://doi.org/10.1007/978-0-387-85820-3\\_11](https://doi.org/10.1007/978-0-387-85820-3_11)

Castells P, Vargas S, Wang J (2011) Novelty and Diversity Metrics for Recommender Systems: Choice, Discovery and Relevance. In: *DDR-2011: International Workshop on Diversity in Document Retrieval at the ECIR 2011: the 33rd European Conference on Information Retrieval*, p 8, URL <http://hdl.handle.net/10486/666094>

Center CINI (2018) *The 41st Statistical Report on Internet Development in China*. URL <http://www.cac.gov.cn/files/pdf/cnnic/CNNIC41.pdf>

Chen L, Wang N, Yang Y, Yang K, Yuan Q (2019a) User Validation of Recommendation Serendipity Metrics. In: *arXiv:1906.11431 [Cs]*, 1906.11431

Chen L, Yang Y, Wang N, Yang K, Yuan Q (2019b) How Serendipity Improves User Satisfaction with Recommendations? A Large-Scale User Evaluation. In: *The World Wide Web Conference, ACM, New York, NY, USA, WWW '19*, pp 240–250, DOI 10.1145/3308558.3313469, URL <http://doi.acm.org/10.1145/3308558.3313469>

- Chiu YS, Lin KH, Chen JS (2011) A Social Network-based Serendipity Recommender System. In: 2011 International Symposium on Intelligent Signal Processing and Communications Systems (ISPACS), IEEE, pp 1–5
- Cooper-Martin E (1991) Consumers and Movies: Some Findings on Experiential Products. *ACR North American Advances NA-18*
- Costa PT, McCrae RR (1992) NEO PI-R Professional Manual. *Psychological Assessment Resources* 396
- De Gemmis M, Lops P, Semeraro G, Musto C (2015) An Investigation on the Serendipity Problem in Recommender Systems. *Information Processing and Management* 51(5):695–717, DOI 10.1016/j.ipm.2015.06.008, URL <http://dx.doi.org/10.1016/j.ipm.2015.06.008>
- Farrar DE, Glauber RR (1967) Multicollinearity in Regression Analysis: The Problem Revisited. *The Review of Economics and Statistics* 49(1):92–107
- Ford GT, Smith DB, Swasy JL (1988) An Empirical Test of the Search, Experience and Credence Attributes Framework NA-15, URL <http://acrwebsite.org/volumes/6817/volumes/v15/NA-15>
- Gosling SD, Rentfrow PJ, Swann WB (2003) A Very Brief Measure of the Big-Five Personality Domains. *Journal of Research in Personality* 37(6):504–528, DOI 10.1016/S0092-6566(03)00046-1, URL <http://linkinghub.elsevier.com/retrieve/pii/S0092656603000461>
- Hildebrand DK, Laing JD, Rosenthal H (1977) *Analysis of Ordinal Data*. 8, Sage

Hu R, Pu P (2013) Exploring Relations between Personality and User Rating Behaviors. In: EMPIRE 1st Workshop on “Emotions and Personality in Personalized Services”, p 12

Huang J, Ding S, Wang H, Liu T (2018) Learning to Recommend Related Entities with Serendipity for Web Search Users. *ACM Transactions on Asian and Low-Resource Language Information Processing* 17(3):1–22, DOI 10.1145/3185663, URL <https://doi.org/10.1145/3185663>

Iacobucci D, Posavac SS, Kardes FR, Schneider MJ, Popovich DL (2015) The Median Split: Robust, Refined, and Revived. *Journal of Consumer Psychology* 25(4):690–704, DOI <https://doi.org/10.1016/j.jcps.2015.06.014>

Jung CG (1983) *Memories, Dreams, Reflections*. London: Fontana Press

Kaminskas M, Bridge D (2014) Measuring Surprise in Recommender Systems. In: *Proceedings of the ACM RecSys Workshop on Recommender Systems Evaluation: Dimensions and Design (Workshop Programme of the 8th ACM Conference on Recommender Systems)*

Kaminskas M, Bridge D (2016) Diversity, Serendipity, Novelty, and Coverage: A Survey and Empirical Analysis of Beyond-Accuracy Objectives in Recommender Systems. *ACM Transactions on Interactive Intelligent Systems* 7(1):2:1–2:42, DOI 10.1145/2926720, URL <http://doi.acm.org/10.1145/2926720>

Kashdan TB, Gallagher MW, Silvia PJ, Winterstein BP, Breen WE, Terhar D, Steger MF (2009) The Curiosity and Exploration Inventory-II: Development, Factor Structure, and Psychometrics. *Journal of Research in Personality* 43(6):987–998, DOI 10.1016/j.jrp.2009.04.011

- Kawamae N (2010) Serendipitous Recommendations via Innovators. In: Proceeding of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '10, ACM Press, p 218, DOI 10.1145/1835449.1835487, URL <http://portal.acm.org/citation.cfm?doid=1835449.1835487>
- Kawamae N, Sakano H, Yamada T (2009) Personalized Recommendation Based on the Personal Innovator Degree. In: Proceedings of the Third ACM Conference on Recommender Systems - RecSys '09, ACM Press, pp 329–332, DOI 10.1145/1639714.1639780, URL <http://portal.acm.org/citation.cfm?doid=1639714.1639780>
- Khoshahval S, Farnaghi M, Taleai M, Mansourian A (2018) A Personalized Location-Based and Serendipity-Oriented Point of Interest Recommender Assistant Based on Behavioral Patterns. In: Geospatial Technologies for All, Springer International Publishing, pp 271–289, DOI 10.1007/978-3-319-78208-9\_14, URL [http://link.springer.com/10.1007/978-3-319-78208-9\\_14](http://link.springer.com/10.1007/978-3-319-78208-9_14)
- Kotkov D, Wang S, Veijalainen J (2016) A Survey of Serendipity in Recommender Systems. Knowledge-Based Systems 111:180–192, DOI 10.1016/j.knosys.2016.08.014, URL <https://doi.org/10.1016/j.knosys.2016.08.014>
- Kotkov D, Konstan JA, Zhao Q, Veijalainen J (2018a) Investigating Serendipity in Recommender Systems Based on Real User Feedback. In: Proceedings of the 33rd Annual ACM Symposium on Applied Computing, ACM, New York, NY, USA, SAC '18, pp 1341–1350, DOI 10.1145/3167132.3167276, URL <http://doi.acm.org/10.1145/3167132.3167276>

Kotkov D, Veijalainen J, Wang S (2018b) How Does Serendipity Affect Diversity in Recommender Systems? A Serendipity-Oriented Greedy Algorithm. *Computing* pp 1–19, DOI 10.1007/s00607-018-0687-5, URL <https://doi.org/10.1007/s00607-018-0687-5>

Li X, Jiang W, Chen W, Wu J, Wang G (2019) HAES: A New Hybrid Approach for Movie Recommendation with Elastic Serendipity. In: *Proceedings of the 28th ACM International Conference on Information and Knowledge Management - CIKM '19*, ACM Press, pp 1503–1512, DOI 10.1145/3357384.3357868, URL <http://dl.acm.org/citation.cfm?doid=3357384.3357868>

Li X, Jiang W, Chen W, Wu J, Wang G, Li K (2020) Directional and Explainable Serendipity Recommendation. In: *Proceedings of The Web Conference 2020*, ACM, pp 122–132, DOI 10.1145/3366423.3380100, URL <https://dl.acm.org/doi/10.1145/3366423.3380100>

Lu Q, Chen T, Zhang W, Yang D, Yu Y (2012a) Serendipitous Personalized Ranking for Top-N Recommendation. In: *2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, IEEE, vol 1, pp 258–265

Lu Q, Chen T, Zhang W, Yang D, Yu Y (2012b) Serendipitous Personalized Ranking for Top-N Recommendation. In: *2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, IEEE, Macau, China, pp 258–265, DOI 10.1109/WI-IAT.2012.135

Maccatrozzo V, Terstall M, Aroyo L, Schreiber G (2017) SIRUP: Serendipity In Recommendations via User Perceptions. In: *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, ACM, New York,

NY, USA, IUI '17, pp 35–44, DOI 10.1145/3025171.3025185, URL <http://doi.acm.org/10.1145/3025171.3025185>

Maslow AH (1943) A Theory of Human Motivation. *Psychological Review* 50(4):370

Matt C, Hess T, Benlian A, Weiß C (2014) Escaping from the Filter Bubble? The Effects of Novelty and Serendipity on Users' Evaluations of Online Recommendations. Publications of Darmstadt Technical University, Institute for Business Studies (BWL) 66193, Darmstadt Technical University, Department of Business Administration, Economics and Law, Institute for Business Studies (BWL), URL <https://ideas.repec.org/p/dar/wpaper/66193.html>

McNee SM, Riedl J, Konstan JA (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, New York, NY, USA, CHI EA '06, pp 1097–1101, DOI 10.1145/1125451.1125659, URL <http://doi.acm.org/10.1145/1125451.1125659>

Menk A, Sebastia L, Ferreira R (2017) CURUMIM: A Serendipitous Recommender System for Tourism Based on Human Curiosity. In: 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI), pp 788–795, DOI 10.1109/ICTAI.2017.00124

Mobile A (2018) 2017 Online Shopping App Market Research Report. URL <https://community.jiguang.cn/article/246360>

Nakatsuji M, Fujiwara Y, Tanaka A, Uchiyama T, Fujimura K, Ishida T (2010) Classical Music for Rock Fans?: Novel Recommendations for Expanding

User Interests. In: Proceedings of the 19th ACM International Conference on Information and Knowledge Management, ACM, New York, NY, USA, CIKM '10, pp 949–958, DOI 10.1145/1871437.1871558, URL <http://doi.acm.org/10.1145/1871437.1871558>

Niu X, Abbas F (2017) A Framework for Computational Serendipity. In: Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization - UMAP '17, ACM Press, pp 360–363, DOI 10.1145/3099023.3099097, URL <http://dl.acm.org/citation.cfm?doid=3099023.3099097>

Oliver MB, Raney AA (2011) Entertainment as Pleasurable and Meaningful: Identifying Hedonic and Eudaimonic Motivations for Entertainment Consumption. *Journal of Communication* 61(5):984–1004, DOI 10.1111/j.1460-2466.2011.01585.x

Onuma K, Tong H, Faloutsos C (2009) TANGENT: A Novel, “Surprise-me”, Recommendation Algorithm. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, pp 657–666

Pandey G, Kotkov D, Semenov A (2018) Recommending Serendipitous Items using Transfer Learning. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management, ACM, pp 1771–1774

Pantouvakis A (2010) The Relative Importance of Service Features in Explaining Customer Satisfaction: A Comparison of Measurement Models. *Managing Service Quality: An International Journal* 20(4):366–387, DOI 10.1108/09604521011057496

- Schedl M, Hauger D, Farrahi K, Tkalčič M (2015) On the Influence of User Characteristics on Music Recommendation Algorithms. In: Hanbury A, Kazai G, Rauber A, Fuhr N (eds) *Advances in Information Retrieval*, Springer International Publishing, Lecture Notes in Computer Science, pp 339–345, DOI 10.1007/978-3-319-16354-3\_37
- Shen T, Jia J, Li Y, Ma Y, Bu Y, Wang H, Chen B, Chua TS, Hall W (2020) PEIA: Personality and Emotion Integrated Attentive Model for Music Recommendation on Social Media Platforms. In: *The Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20)*, p 8
- Tuval N (2019) Exploring the Potential of the Resolving Sets Model for Introducing Serendipity to Recommender Systems. In: *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*, Association for Computing Machinery, UMAP '19, pp 353–356, DOI 10.1145/3320435.3323467, URL <https://doi.org/10.1145/3320435.3323467>
- Walpole H (1960) To Mann, Monday 18 January 1754. In: *Horace Walpole's Correspondence*, Yale University Press, pp 407–411
- Wang CD, Deng ZH, Lai JH, Yu PS (2018) Serendipitous Recommendation in E-Commerce Using Innovator-Based Collaborative Filtering. *IEEE Transactions on Cybernetics* 49(7):2678–2692, DOI 10.1109/TCYB.2018.2841924
- Wang N, Chen L, Yang Y (2020) 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*, Association for Computing Machinery, UMAP '20, pp 266–274, DOI 10.1145/3340631.3394863, URL <https://doi.org/10.1145/3340631>.

[3394863](#)

Wikipedia Contributors (2021) Window Shopping — Wikipedia, The Free Encyclopedia. URL [https://en.wikipedia.org/w/index.php?title=Window\\_shopping&oldid=1042448346](https://en.wikipedia.org/w/index.php?title=Window_shopping&oldid=1042448346), [Online; accessed 19-September-2021]

Wikipedia Contributors (2022) Mann–Whitney U test — Wikipedia, The Free Encyclopedia. URL [https://en.wikipedia.org/w/index.php?title=Mann%E2%80%93Whitney\\_U\\_test&oldid=1071063635](https://en.wikipedia.org/w/index.php?title=Mann%E2%80%93Whitney_U_test&oldid=1071063635), [Online; accessed 30-March-2022]

Wu W, Chen L (2015) Implicit Acquisition of User Personality for Augmenting Movie Recommendations. In: Ricci F, Bontcheva K, Conlan O, Lawless S (eds) *User Modeling, Adaptation and Personalization*, Springer International Publishing, Lecture Notes in Computer Science, pp 302–314, DOI 10.1007/978-3-319-20267-9\_25

Zhang YC, Séaghdha DO, Quercia D, Jambor T (2012) Auralist: Introducing Serendipity into Music Recommendation. In: *Proceedings of the 5th ACM International Conference on Web Search and Data Mining*, ACM, New York, NY, USA, WSDM '12, pp 13–22, DOI 10.1145/2124295.2124300, URL <http://doi.acm.org/10.1145/2124295.2124300>

Zhao P, Lee DL (2016) How Much Novelty is Relevant?: It Depends on Your Curiosity. In: *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '16*, ACM Press, pp 315–324, DOI 10.1145/2911451.2911488, URL <http://dl.acm.org/citation.cfm?doid=2911451.2911488>

Zheng Q, Chan CK, Ip HHS (2015) An Unexpectedness-Augmented Utility Model for Making Serendipitous Recommendation. In: Perner P (ed) *Advances in Data Mining: Applications and Theoretical Aspects*, vol 9165, Springer International Publishing, Cham, pp 216–230, DOI 10.1007/978-3-319-20910-4\_16