The Impacts of Item Features and User Characteristics on Users' Perceived Serendipity of Recommendations

Ningxia Wang
Hong Kong Baptist University
Hong Kong, China
nxwang@comp.hkbu.edu.hk

Li Chen
Hong Kong Baptist University
Hong Kong, China
lichen@comp.hkbu.edu.hk

Yonghua Yang
Alibaba Group
China
huazai.yyh@alibaba-inc.com

ABSTRACT

Serendipity-oriented recommender systems have increasingly been recognized as useful to overcome the "filter bubble" problem of accuracy-oriented recommenders, by recommending *unexpected* and *relevant* items to users. However, most of existing systems are based on researchers' assumptions about the effect of item features on serendipity, but less from users' perspective to study what item features and even user characteristics might affect their perceived serendipity. In this paper, we have attempted to fill in this vacancy based on results of a large-scale user survey (involving over 10,000 users). We have analyzed the correlation between different types of features (i.e., numerical and categorical) with user perceptions, and furthermore identified the interaction effect from user characteristics (such as personality traits and curiosity). We finally discuss the implications of our work to augment the effectiveness of current serendipity-oriented recommender systems.

KEYWORDS

Recommender systems, serendipity, item features, user personality, curiosity, user survey

ACM Reference Format:
Ningxia Wang, Li Chen, and Yonghua Yang. 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 (UMAP '20), July 14–17, 2020, Genoa, Italy*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3340631.3394863

1 INTRODUCTION

Serving as popular tools of filtering massive information and helping users make decisions, recommender systems (RSs) bring huge benefits to both users and suppliers. Traditional RSs (like collaborative based RSs) have mainly aimed to maximize the accuracy of item prediction, but are likely to trap users into a "filter bubble"[22, 23], since the recommendations that are too similar to the user's previous preference may stop her/him from exploring different and new items. Given this limitation, *serendipity-oriented RSs* have been proposed, which, instead of stressing purely on accuracy or novelty [3, 13], are targeted to balance both relevance and surprise. A recent study showed that increased serendipity can significantly lead to higher user satisfaction with the recommendation [4].

The basic idea behind existing serendipity-oriented RSs is to capture item features that may satisfy users' needs for pleasant surprises. For example, it is assumed that unpopular items will bring more uncertainty and hence be out of the user's expectation, based on which the importance of unpopular items is increased when generating serendipitous recommendations [20]. In another work, it is assumed that the more dissimilar an item is to the user's profile (i.e., her/his previously visited items), the more surprising it is [18, 34]. Users' personal characteristics have also been considered in some work. For instance, highly curious users are assumed to be more likely to accept novel items, while users with low curiosity may prefer to receive accuracy-oriented recommendations [21, 24, 27].

However, to the best of our knowledge, little work stands from end-users' perceptive to study what item features and/or user characteristics would in practice affect their perceived serendipity of the recommendation, which may impair the applicability of existing serendipity-oriented RSs in real-life situations [17]. Therefore, in this work, we have been engaged in answering the following three research questions through a user survey:

- *RQ1*: What item features can influence users' perceived serendipity?
- *RQ2*: What user characteristics can affect their perceived serendipity, and furthermore how would they interact with item features to take the effect?
- *RQ3*: What inspirations can this work bring to the design of serendipity-oriented recommender systems?

We concretely conducted a large-scale user survey (involving over 10,000 users) on a commercial platform to collect users' feedback on recommendations, as well as their demographic information (e.g., age, gender, personality, and curiosity). There are several interesting findings from statistical analyses: 1). Lower item popularity, smaller time difference (the time distance between the current recommendation and the latest item of the same category in the user's profile), smaller category difference (the difference in terms of category level in the item taxonomy), and shorter taxonomic distance (the distance between two items' leaf categories over the taxonomy) are more significantly related to users' perceived higher serendipity. 2). Four association rules mined from items' categorical features reveal users' behavior patterns. 3). Higher-age and/or male users, or users with higher curiosity, higher openness to experience, higher conscientiousness, higher extraversion, higher neuroticism, or lower agreeableness are easier to feel the recommendation serendipitous. 4). Some item features do not interact

ACM ISBN 978-1-4503-6861-2/20/07…$15.00
https://doi.org/10.1145/3340631.3394863

with user characteristics, such as time difference based on long-term profile, while some features significantly interact with, such as category difference.

In short, there are five major contributions of this work: 1). We have tested a number of item features and user characteristics on users' perceived serendipity of the recommendation through a large-scale user survey. 2). For item features, we have particularly taken into account product types (according to SEC classification [9, 26] and types of consumer products [8]). 3). For user characteristics, we have analyzed the respective influences of user personality traits and curiosity. 4). We have further identified the interaction effects between item features and user characteristics. 5). Several design implications are derived from our experimental findings.

The remainder is organized as follows. We first introduce the related work on serendipity-oriented RSs, with focus on item features and user characteristics they have considered (Section 2). We then present our experimental methodology (Section 3), followed by results analysis (Section 4). Near the end, we discuss the implications of our work (Section 5) and future directions (Section 6).

2 RELATED WORK

The original definition of serendipity is "[...] *making discoveries, by accidents and sagacity, of things which they were not in quest for* [...]" [2, 21, 30]. In RSs, it has been emphasized to break through the barriers of traditional accuracy-oriented approaches, for triggering users' positive emotion to unexpected recommendations. [19] divided serendipity-oriented recommending methods into three categories, i.e., reranking, modification, and novel algorithms. In this section, we mainly summarize **explicit item features** and **user characteristics** that related methods have considered. Note that the review does not include those algorithms based on implicit features via graph-based [7, 28] or neural networks approaches [29].

One of the most commonly considered item features is *item popularity*. Unpopular items are likely to be unknown and hence assumed to be unexpected than popular ones. For example, [20] revised the optimization function of matrix factorization by adding a popularity term to underestimate the importance of popular items. Another frequently used item feature is *item (dis)similarity to the user profile*. The more different/dissimilar an item is from the set of items visited by the user before, the more unexpected it is assumed to be [34]. In [34], the authors combined both item unpopularity and dissimilarity to define unexpectedness and proposed an unexpectedness-augmented PureSVD latent factor model. [18] proposed a hybrid reranking algorithm that also considers both unpopularity and dissimilarity. Both works defined popularity as the number of user visits, and dissimilarity to the user profile as the average of pairwise item dissimilarities. Besides, [25] proposed a class distance metric based on item taxonomy, and defined item novelty as the smallest distance from the target item's class to the classes that the user have lately accessed.

In addition, [15] measured *the estimated purchase time* to predict when the user would purchase an item, and the proposed recommending algorithm was to reduce the time cost for the user to find that item. This time-saving strategy may lead to surprising effect as claimed by the authors. [32] introduced a hybrid rank-interpolation music RS that was aimed at balancing multiple metrics to improve user satisfaction. Specifically, they considered *listener diversity* (the entropy of the artist's listener distribution) and *clustering* (a graph-based measure of how nodes are clustered) to enhance serendipity. [5] proposed a social network based serendipity RS, which considers the *recency* (the item's access time).

As for user characteristics, one of the most commonly considered characteristics is *curiosity* [4], because it is an important premise of users' appetite for novelty in the field of psychology [1, 33]. [21] developed a serendipity model based on curiosity theory, which estimates a user's coping potential via item diversity in her/his profile. [24] proposed to predict user curiosity from her/his generated data in social networking, and then generate serendipitous recommendations being tailored to the user's curiosity value. [27] designed a framework for computational serendipity, which adapts the surprise degree of recommendations to the curiosity curve, so as to trigger the user's curiosity at the appropriate time.

Another type of user characteristic is *innovator* [16], referring to those consumers who are more sensitive to new items. Items purchased by them are assumed to surprise the followers. [31] stated that innovators should have three properties: Having high user activity, strong ability of discovering new items and taking shorter time to find unpopular items, and being unlikely to follow the mainstream. The recommendations to the current user were then retrieved from items that the nearest innovators have interacted with.

Although various item features and user characteristics have been incorporated into serendipity-oriented RSs so far, they are mostly based on researchers' assumptions. To the best of our knowledge, little work has empirically validated their relationships with users' perceived recommendation serendipity. The novelty of our work lies in verifying the effects of major item features and user characteristics on user perception. Moreover, we identify some new item features (e.g., time difference and category difference) that were not discussed in related work but show significant correlations with user perception in our experiment. As for user characteristics, in addition to curiosity, we find that users' demographic characteristics (e.g., age and gender) and personality traits (e.g., openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism) significantly affect users' perceived serendipity.

3 METHODOLOGY

In this section, we first introduce the conducted user survey, followed by the set of item features and user characteristics we have investigated through this experiment.

3.1 User Survey

We conducted a user survey on a popular mobile e-commerce application in China (i.e., *Mobile Taobao*) starting from Dec. 21, 2017. If a user of this app volunteered to join, s/he first answered several questions about her/his demographic background (e.g., age and gender). S/he was also asked to filled out two psychological quizzes (see Table 1): One about her/his curiosity via ten-item Curiosity and Exploration Inventory-II (CEI-II) [14], and the other about personality by Ten-Item Personality Inventory (TIPI) [10]. Note that we tend

Table 1: User answers to survey questions

Question	Mean (Std.)	Median	K-S test
Serendipity: *"The item recommended to me is a pleasant surprise."*	2.650 (1.454)	2.0	0.200***
Curiosity: Curiosity and Exploration Inventory-II (CEI-II) [14]	3.138 (0.819)	3.1	0.035***
Big-Five Personality: Ten-Item Personality Inventory (TIPI) [10]			
-*Openness to Experience*	4.627 (1.286)	4.5	0.093***
-*Conscientiousness*	4.562 (1.452)	4.5	0.089***
-*Extraversion*	4.173 (1.651)	4.0	0.085***
-*Agreeableness*	4.969 (1.069)	5.0	0.130***
-*Neuroticism*	4.260 (1.426)	4.0	0.096***

Note: Questions about serendipity and curiosity were responded on 5-point Likert scale from 1 "strongly disagree" to 5 "strongly agree", and those about personality were on 7-point Likert scale. All questions were accompanied by Chinese translations. $^{***}p < 0.001$ for Kolmogorov-Smirnov test (that means the distribution is not normal).

Table 2: Statistics of user profile

# of valid users in the survey	11,383
# 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
Profile duration	Past 3 months
Average # of items clicked per user	1,802.219 (min = 7, max = 19,043)
Average # of items purchased per user	72.437 (min = 0, max = 1,858)

Note: The **user profile** refers to the set of items clicked and/or purchased by the user in the past three months.

to choose short psychological quizzes for the purpose of avoiding users becoming impatient when they fill out the questionnaire.

Then, the user received a recommended product (with its name, image, short description, and price) as generated by *Mobile Taobao*, and gave her/his immediate feedback on the recommendation in terms of its serendipity (see Table 1). As the incentive, all of the participants were placed in a lottery draw with customized presents as awards given to the winners.

Till March 17, 2018, we received 13,741 users' responses and got their consent to use the data for research. We carefully checked all of the responses in order to filter out invalid answers. For example, if a user did not answer all of the questions, or gave the same rating to two opposite questions, her/his response was deleted. In addition, we only kept the user's first response if s/he joined in the survey more than once. As a result, 11,446 users remained, among whom we further removed some outlier cases (17 cases in which the user clicked either less than 5 or more than 15,000 items before taking the survey, 12 cases where the recommendation's top-level category appeared less than 10 times within all users' profiles, and 34 cases where the recommendation's top-level category is "Uncategorized"). Finally, we have 11,383 users' records (7,769 females), including each one's past three months' historical data (i.e., the items s/he had clicked or purchased). Table 1 lists the descriptive analysis of their answers to our questions, and Table 2 shows the statistics of their profiles.

Besides, an item taxonomy was obtained from *Mobile Taobao* to show items' hierarchical classification (see Figure 1). In total, there are 20 top-level categories (e.g., "Clothes", "Toys", "Home appliances", etc.), and 91 sub-categories under them. The leaf category at the bottom of each path is the direct category that a particular item belongs to.

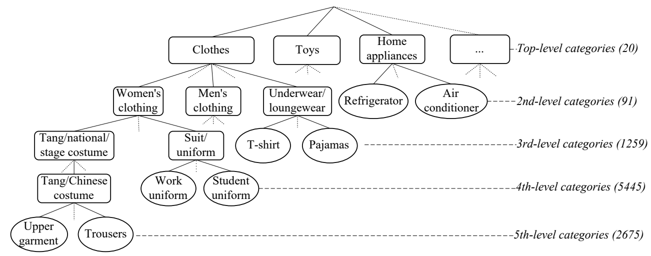

Figure 1: The item taxonomy (totally 9,490 categories after pre-cleaning) from *Mobile Taobao*.

Table 3: Summary of numerical item features

Feature name	Short description
Non-personalized item features	
Popularity	pop, a binary function based on a set of the most popular items.
Personalized item features	
Time difference	$TimeDiff$, the time distance from the recommendation r to the latest item in the user's profile that belongs to the same category of r.
Category difference	At which category level that two items differ from over the taxonomy. Three variations: The recent category difference ($CateDiff_{rec}$), the average category difference ($CateDiff_{avg}$), and the minimal category difference ($CateDiff_{min}$).
Taxonomic distance	The minimum number of hops between two items' leaf categories over the taxonomy. Three variations: The recent taxonomic distance ($TaxoDist_{rec}$), the average taxonomic distance ($TaxoDist_{avg}$), and the minimal taxonomic distance ($TaxoDist_{min}$).

3.2 Item Features

With the collected user data, we have extracted a number of item features, which can be in general divided into *numerical* and *categorical* types. Numerical features (see Table 3) can further be divided into *non-personalized* and *personalized* features based on whether the feature is independent of the target user's profile or not.

3.2.1 Numerical item features. The numerical feature means that the feature's value can be quantified via an algorithmic function.

(1) *Non-personalized.* The non-personalized feature is independent of the target user's profile. As mentioned in related work, item unpopularity has been typically used to indicate its surprise level [13], for which the **popularity** is usually determined by the total number of users who have visited that item. In our case, we define popularity as a binary function by calling HOT as provided by Mobile Taobao [4]. Specifically, HOT will return the item with the most clicks at the time of calling, so if an item belongs to the set of items ever returned by HOT, its popularity is 1, otherwise 0.

(2) *Personalized.* Personalized item feature is to reveal the relationship between the current recommendation r and those in the target user u's profile P_u (consisting of items that u has previously visited). In our collected user data, each user has two profiles: P_u^{click} (the previously clicked items) and $P_u^{purchase}$ (the previously purchased items). A timestamp t_{ui}

is associated with each item i to indicate when it was visited by the corresponding user. In the following, we introduce feature functions we have proposed, or referred to related work, to utilize this information.

- **Time difference.** This feature can show the time distance between the current recommendation r and the latest one in user profile (P_u^{click} or $P_u^{purchase}$) that belongs to the same category of r (where category refers to that defined in the item taxonomy; see Figure 1). The function is given in Equation (1):

$$TimeDiff(r, P_u) = t_{ur} - t_{ui_{cr}} \quad (1)$$

where t_{ur} is the recommendation r's timestamp, and $t_{ui_{cr}}$ denotes the latest timestamp of u visiting an item i that belongs to the same category of r (i.e., c_r). In more detail, c_r has two variations: c_r^{top} refers to the top-level category of r, and c_r^{leaf} refers to its leaf category.

- **Category difference.**
It measures at which category level the two compared items become different over the item taxonomy. To be more specific, there are in total 5 levels in our taxonomy (as shown in Figure 1), so the largest possible category difference is 6 (when the two items do not share any common category, i.e., being different starting from the top-level category), and the smallest is 1 (when the two items have the same leaf category at the 5-th level). Formally, it can be calculated as $6-|C_r \cap C_i|$, where C_r and C_i respectively denote the category sets of the recommendation r and the item i from the user's profile P_u (where P_u can be P_u^{click} or $P_u^{purchase}$).

There can be three ways of calculating the category difference: a. The *recent category difference* that refers to the category difference between the recommendation r and the recent (latest) item i_{rec} in the user's profile:

$$CateDiff_{rec}(r, P_u) = 6 - |C_r \cap C_{i_{rec}}| \quad (2)$$

b. The *average category difference* between the recommendation r and all items in the user's profile:

$$CateDiff_{avg}(r, P_u) = \frac{1}{|P_u|} \sum_{i \in P_u} (6 - |C_r \cap C_i| \quad (3)$$

c. The *minimal category difference* between the recommendation r and all items in the user's profile:

$$CateDiff_{min}(r, P_u) = min_{i \in P_u}(6 - |C_r \cap C_i|) \quad (4)$$

Considering the average and minimal differences can be sensitive to the duration of the user's profile, we built two more variations that only consider the user's historical data in the past 15 days till the survey, i.e., the user profiles $P_u^{click\_15d}$ and $P_u^{purchase\_15d}$ respectively.

- **Taxonomic distance.** The taxonomic distance between two items i and j is defined as the minimum number of hops jumping from i's leaf category c_i^{leaf} to j's leaf category c_j^{leaf} over the item taxonomy, which is equivalent to the class distance proposed in [25]. Similar to category difference, there are also 3 ways of computing the taxonomic distance: a. The *recent taxonomic distance* between the recommendation and the recent item in the user's profile:

$$TaxoDist_{rec}(r, P_u) = hops(c_r^{leaf}, c_{i_{rec}}^{leaf}) \quad (5)$$

b. The *average taxonomic distance* between r and all items in the user's profile:

$$TaxoDist_{avg}(r, P_u) = \frac{1}{|P_u|} \sum_{i \in P_u} hops(c_r^{leaf}, c_i^{leaf}) \quad (6)$$

c. The minimal taxonomic distance:

$$TaxoDist_{min}(r, P_u) = min_{i \in P_u} hops(c_r^{leaf}, c_i^{leaf}) \quad (7)$$

Still, for the average and minimal distances, two more variations based on the 15-day user profile are considered.

Table 4: General product classification schemes

Scheme	Type	Explanation	Instances
SEC classification [9, 26]	*Search*	Evaluable prior to purchase	Clothing, office stationery
	Experience	Evaluable based on purchase experience	Hairdresser, beauty salon
	Credence	Evaluable based on professional knowledge	Legal services, medical treatment
Consumer product type [8]	*Convenience*	With low involvement effort and risk	Sugar, magazines
	Shopping	Consumers are willing to spend more time in searching and comparing before purchase	Automobiles, clothing
	Specialty	With unique characteristics or brand identification; some users will go to great lengths to get them	Vintage wines, expensive cars
	Unsought	Users do not know about them or do not normally think of buying	Reference books

3.2.2 Categorical item features. Categorical item features are aimed to identify the association between categories of the current recommendation and those appearing in the user profile. Specifically, we apply association rule mining over items' category labels in order to find out the most confident (i.e., frequently occurring) rules for both *serendipitous recommendations* and *non-serendipitous ones.* Moreover, in addition to considering categories defined in the item taxonomy (see Figure 1), we assign some upper-level product type labels to each item according to the general classification scheme [8, 9, 26].

(1) **Taxonomy category.** Each item is assigned three category labels according to the taxonomy (see Figure 1).
    - *Top-level category* (e.g., "Clothes", "Home applicances", etc.) that indicates a coarse-granularity categorization.
    - *2nd-level category* (e.g., "Men's clothing" under "Clothes"), which is the immediate sub-category under the top-level category. Because every item has the 2nd-level category in our case, we use it to represent the moderate granularity.
    - *Leaf category* (e.g., "Upper garment" and "Trousers" at the 5th level), which is the category an item directly belongs to, so it is of the finest granularity.

(2) **Product type.** There are different product classifications as developed in the fields of marketing and economics [8], among which we selected two of the most commonly used schemes (see Table 4): *SEC classification* from the aspect of whether users can acquire enough information to support

their evaluation of the product [9, 26], and *consumer product type* in consideration of the effort required by consumers and the potential risk to purchase a product [8]. In total, there are 12 (= 3 x 4) label combinations. We implemented a mapping algorithm that automatically assigns product type labels to each item by linking its 2nd-level category (given its general price range, purchase frequency, brand recognition, durability, etc.) to the two classification schemes.

3.3 User Characteristics

In our survey, we also acquired users' basic demographic properties and psychological traits.

3.1.1 Demographic characteristics. As shown in Table 2, our users are distributed among different *age* groups, i.e., 18-20, 20-30, 30-40, 40-50, 50-60, and above 60 years old. Another property is *gender*. There are about 2/3 female participants and 1/3 males. In terms of other demographic features (e.g., nationality and job domain), we did not include them in this work, because most of participants are Chinese, and the variety of job domains makes the analysis complicated.

3.1.2 Psychological characteristics. Inspired by related work, we obtained users' curiosity in order to analyze its effect on their serendipity perception. In addition, given that curiosity has been found significantly correlated with users' personality traits (such as openness to experience, conscientiousness, extraversion, and neuroticism [14]), we also included the Big-Five factor model [10] to identify the five major personality traits' respective influences.

- **Curiosity.** As mentioned before, we adopted a popularly used curiosity quiz, i.e., Curiosity and Exploration Inventory-II (CEI-II) [14], to measure whether the user has a strong desire for new knowledge or experience in general. Indeed, curiosity has been widely regarded as an important antecedent of users' appetite for novelty in the field of psychology [1]. It can greatly affect the level of pleasure a user may experience when s/he explores new and surprising things.
- **Big-Five Personality.** The Big-Five factor model (also known as the OCEAN model) is a popularly used taxonomy to define a person's personality [10] from the following five aspects:
  - *Openness to Experience* indicates the person's facets like 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" [12]. People with high extraversion tend to be enthusiastic, outgoing, and talkative.
  - *Agreeableness* implies facets like trust, altruism, and tender-mindedness. People scoring high on this trait are empathetic and willing to cooperate.
  - *Neuroticism* is about anxiety, hostility, depression, self-consciousness, impulsiveness, and vulnerability [6]. People who score higher on neuroticism are more likely to be moody, feel stressful, and have difficulty in delaying gratification.

Table 5: Results of independent-samples Mann-Whitney U tests regarding numerical item features

Feature	Variations	Mann-Whitney U	Mean of the low (high) serendipity group
$TimeDiff$	P_u^{click}, c_r^{top}	14,350,794***	5.028 (**3.675**)
	P_u^{click}, c_r^{leaf}	5,887,957***	12.121 (**9.241**)
	$P_u^{purchase}, c_r^{top}$	9,600,291***	16.241 (**14.761**)
	$P_u^{purchase}, c_r^{leaf}$	1,073,343***	22.706 (**18.189**)
$CateDiff_{rec}$	P_u^{click}	14,925,459***	5.19 (**5.00**)
	$P_u^{purchase}$	15,120,992***	5.50 (**5.38**)
$CateDiff_{avg}$	P_u^{click}	15,898,945.500	5.58 (5.59)
	$P_u^{purchase}$	15,737,353.500	5.68 (5.68)
	$P_u^{click\_15d}$	15,185,636***	5.50 (**5.47**)
	$P_u^{purchase\_15d}$	13,622,498***	5.62 (**5.58**)
$CateDiff_{min}$	P_u^{click}	14,643,922.500***	3.22 (**3.06**)
	$P_u^{purchase}$	15,266,937.500***	4.16 (**4.08**)
	$P_u^{click\_15d}$	14,036,685***	3.75 (**3.47**)
	$P_u^{purchase\_15d}$	13,399,036***	4.79 (**4.65**)
$TaxoDist_{rec}$	P_u^{click}	15,235,757***	5.22 (**4.90**)
	$P_u^{purchase}$	15,421,617*	5.97 (**5.78**)
$TaxoDist_{avg}$	P_u^{click}	16,667,286.500***	**6.08** (6.16)
	$P_u^{purchase}$	16,206,497.000*	**6.41** (6.45)
	$P_u^{click\_15d}$	28,954,051	5.90 (6.17)
	$P_u^{purchase\_15d}$	14,036,586	6.24 (6.50)
$TaxoDist_{min}$	P_u^{click}	15,057,039.000***	1.32 (**1.08**)
	$P_u^{purchase}$	15,496,016.000	3.09 (3.01)
	$P_u^{click\_15d}$	14,452,246***	2.29 (**1.84**)
	$P_u^{purchase\_15d}$	13,658,742***	4.27 (**4.09**)

Note: $^{***}p < 0.001$, $^{**}p < 0.01$ and, $^{*}p < 0.05$ for Mann-Whitney U test.

We concretely employed Ten-Item Personality Inventory (TIPI) [10] to measure each participant's personality.

4 RESULTS ANALYSIS

We first analyzed the respective effects of item features and user characteristics on users' perceived recommendation serendipity (see results in Sections 4.1 and 4.2), and then investigated their interaction effects (see results in Section 4.3).

4.1 Impact of Item Features

We transformed users' answers to the serendipity perception question into two groups through median split method [11]. It turns out that there are 5,389 users' responses (47.3%) in the *high* serendipity group (rating higher than 2, i.e., the median value), and 5,994 (52.7%) in the *low* serendipity group.

4.1.1 Numerical item features. For numerical item features, we chose statistical analysis methods to compare distributions of feature values between the two serendipity groups (high and low). Formally, for binary features like popularity, we calculated correlation, and for non-binary values, we conducted the non-parametric test Mann-Whitney U test because the values are not normally distributed (Kolmogorov-Smirnov test $p < 0.05$).

(1) ***Non-personalized numerical item feature.*** Pearson Chi-Square was computed to see the correlation between recommendations' popularity and users' perceived serendipity, whose result is 198.040 ($df = 1$) with asymptotic significance (2-sided) $p < 0.001$. This indicates that there is a significant

relationship between the two variables. We then calculated *Phi* correlation, for which the result is -0.132 ($p < 0.001$). It hence reveals a significantly negative correlation, suggesting that less popular items are more likely to be perceived as serendipitous by users.

(2) ***Personalized numerical item features***

- *Time difference.* Results of Mann-Whitney U tests on all of the four variations related to the time difference feature are significant ($p < 0.001$, see Table 5). It hence indicates that no matter which user profile is considered (clicked or purchased), and which category is referred to (top-level category or leaf category), there are significant differences between high and low serendipity groups regarding the time distance from the current recommendation to the latest one of the same category in user profile. More specifically, as mean values show, the average time differences in the high serendipity group are always shorter, inferring that if a recommended item is in the same category of a more recent item that the user has visited, it might be more serendipitous. This finding contradicts our common assumption that an unexpected item would be likely from a category that the user has not lately accessed.
- *Category difference.* It shows that the category difference in the high serendipity group is smaller than that in the low serendipity group, which is significant for three variations (i.e., *recent*, *average*, and *minimal*). Note that for the *average* variation, the difference is only significant when a short-term user profile (with 15-day length) is considered. These results suggest that the recommendation perceived as serendipitous by users is not necessarily of bigger category difference from those the user has visited. In some cases, even the item is from the same 5-th level leaf category (in which case the category difference is 1), it might still be of serendipity.
- *Taxonomic distance.* Taxonomic distance also takes into account the item's category over the taxonomy, but it primarily calculates the hops it takes from one item's leaf category to another item's, rather than considering the category's layer (level) in the taxonomy. As it is shown in Table 5, the difference between the high and low serendipity groups is more significant for click profile (P_u^{click}). Moreover, in the *average* condition, the mean distance is significantly longer in the high serendipity group, implying that the average distance over the taxonomy between a serendipitous recommendation and the user profile is larger (i.e., traversing more nodes) than a non-serendipitous one; while in the *recent* and *minimal* conditions, it is still shorter in the high serendipity group.

4.1.2 Categorical item features.

For categorical features, we mined the association rules in form of a conditional statement, i.e., *{A, B, ... }* $\land$ *X* $\longrightarrow$ *High/Low serendipity*, meaning that if a user ever purchased items belonging to A (and items belonging to B, ...), there is a high probability that s/he will perceive an item from X category with high/low serendipity. Motivated by the above findings, we fixed the user profile length to 15 days instead of 3 months, and mainly considered the purchase profile $P_u^{purchase}$. Specifically, we run the Apriori algorithm[1] over all users' transactions (the threshold of $support$ is set as 1% and that of $confidence$ is 50%). Each transaction contains the category labels assigned to items purchased by that user (see Section 3.2.2) as antecedents and the *high* or *low* serendipity level of her/his received recommendation as the consequence. Finally we got respectively 324, 375, 313 and 9 rules for the *top-level category*, the *2nd-level category*, the *leaf category*, and the *product type*.

We further filtered out some rules with the following two criteria: 1). The same rule appears in both the *high* and *low* serendipity groups. 2). This rule's antecedent is part of the antecedent of another rule. Below we list four rules with the largest confidence value in respect of each categorical feature:

- Rule 1: As for the *top-level category*, *{"Small appliances", "Jewellery / Watches / Eyewear"}* $\land$ *"Clothes"* $\longrightarrow$ *Low serendipity* ($confidence = 68.25\%$, $support = 2.48\%$), indicating that users, who have purchased products from "Small appliances" and "Jewellery / Watches / Eyewear" categories, tend to perceive the recommendation from category "Clothes" with low serendipity.
- Rule 2: As for the *2nd-level category*, *{"Snacks / Nuts / Specialty", "Makeup / Perfume / Beauty Tools"}* $\land$ *"Women's clothing"* $\longrightarrow$ *Low serendipity* ($confidence = 73.99\%$, $support = 1.23\%$), which infers that users, who have purchased products from "Snacks / Nuts / Specialty" and "Makeup / Perfume / Beauty Tools" categories, tend to perceive a recommendation from "Women's clothing" with low serendipity.
- Rule 3: As for the *leaf category*, *"Sweatshirt / Fleece" / "Sweater" / "Cotton-padded clothes"* $\longrightarrow$ *Low serendipity* ($confidence = 59.46\%, 55.81\%, 55.50\%$ respectively, $support = 1.78\%, 2.07\%$ and 1.98% respectively), inferring that users tend to perceive a recommendation from "Sweatshirt / Fleece", "Sweater" or "Cotton-padded clothes" categories with low serendipity. Note that this rule is independent of the user's purchase profile.
- Rule 4: As for the *product type*, *{"'Experience & Shopping", "Search Convenience", "Search & Specialty"}* $\land$ *"Search & Shopping"* $\longrightarrow$ *Low serendipity* ($confidence = 64.59\%$, $support = 2.48\%$), suggesting that users, who have purchased items of "Experience & Shopping", "Search Convenience" and "Search & Specialty" types, tend to perceive a recommendation of "Search & Shopping" type with low serendipity.

4.2 Impact of User Characteristics

In this part of analysis, we grouped users by their characteristics (e.g., female vs. male for gender analysis) and then compared different groups in terms of their serendipity scores on the 5-point Likert scale.

4.2.1 Demographic characteristics. For *age*, we used Kruskal-Wallis 1-way ANOVA test to perform a multi-group comparison. The result is significant with $p < 0.001$ (as adjusted by the Bonferroni correction). It shows that higher-age users are easier to feel serendipitous. Pairwise comparisons in-depth show that the serendipity levels perceived by users in age groups 18-20, 20-30,

[1]Through the open-source data mining toolbox Weka (https://www.cs.waikato.ac.nz/ml/weka/).

Table 6: Results of independent-samples Mann-Whitney U test regarding psychological traits

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

Note: $^{***}p < 0.001$ for Mann-Whitney U test.

and 30-40 years old are significantly different from those of other three age groups, but there are no significant differences among groups of 40-50, 50-60, and above 60 years old.

For *gender*, non-parametric Mann-Whitney U test was used for between-group comparison, by which the results indicate that male users gave significantly higher serendipity scores on items than female users (mean= 2.79 vs. = 2.59, $p < 0.001$).

4.2.2 Psychological traits. Median split method was used to divide users into two groups in respect of each psychological trait (e.g., high curiosity vs. low). Table 6 shows the number of users in each group and Mann-Whitney U test results. It can be seen that there is a significant difference regarding each trait. Specifically, people with higher score on *curiosity*, *openness to experience*, *conscientiousness*, *extraversion*, or *neuroticism*, are likely to perceive a recommendation as more serendipitous, than those with low score; while for agreeableness, it is in the opposite way (i.e., people of low agreeableness are more inclined to feel the item serendipitous).

Some of the results are consistent with findings of related work, which show that highly curious users are more sensitive to novel and unexpected items [19], and curiosity is significantly correlated with the three personality traits, i.e., openness to experience, conscientiousness, and extraversion [14]. However, there are two new observations in our results: 1). Neuroticism takes a positive effect rather than a negative effect in [14]; 2). agreeableness is significantly negatively related to users' serendipity perception.

4.3 Interaction Effect

In order to know if the effect of item features on serendipity perception varies from person to person, we further analyzed the interaction effect between item features and user characteristics. To this end, we computed a binary logistic regression for each pair of item feature and user characteristic:

$$ln(\frac{ser}{1-ser}) = \beta_0 + \beta_1 \cdot F + \beta_2 \cdot C + \beta_3 \cdot (F \times C) \tag{8}$$

where ser is the serendipity level (high/low), F denotes the item feature (e.g., popularity) and C is one of user characteristics (e.g., curiosity). $F \times C$ is hence the interaction term, for which if the coefficient is significant ($p < 0.05$), it indicates a significant interaction effect, meaning that the effect of the item feature is not the same for different user groups as divided by C. Because we have 20 feature functions (see Table 5) and 8 user characteristics, there are in total 160 logistic regression models.

There are several interesting findings (see Table 7): 1). Five item features do not interact with user characteristics, including click profile based time difference (i.e., $TimeDiff(P_u^{click}, c_r^{top})$ and $TimeDiff(P_u^{click}, c_r^{leaf})$), long-term click profile based minimal category difference ($CateDiff_{min}(P_u^{click})$), long-term click profile based minimal taxonomic distance ($TaxoDist_{min}(P_u^{click})$), and short-term purchase profile based minimal taxonomic distance ($TaxoDist_{min}(P_u^{purchase\_15d})$). 2). Popularity significantly interacts with age. Concretely, unpopular items are 2.012 times more likely to be perceived as serendipitous by users aged 18-20 (1.98 and 1.78 times respectively by users aged 20-30 and 30-40), inferring that young people are more sensitive to item popularity. 3). Users with low curiosity, openness to experience, extraversion, or neuroticism are more sensitive to long-term purchase profile based recent category difference ($CateDiff_{rec}(p_u^{purchase})$) and short-term purchase profile based minimal category difference ($CateDiff_{min}$ $(p_u^{purchase\_15d})$). 4). Average taxonomic distance is affected by user characteristics in an opposite way, i.e., high openness to experience users and high neuroticism users are more sensitive to average taxonomic distance.

Table 7: Results of interaction effects between item features and user characteristics

Item feature	Variation	Age	Q	O	C	E	A	N
$Popularity$		L						
$TimeDiff$	$P_u^{purchase}, c_r^{top}$				H			
	$P_u^{purchase}, c_r^{leaf}$					L		
$CateDiff_{rec}$	P_u^{click}						L	
	$P_u^{purchase}$		L	L		L		L
$CateDiff_{avg}$	$P_u^{click\_15d}$						L	
	$P_u^{purchase\_15d}$			L	L	L	H	L
$CateDiff_{min}$	$P_u^{purchase}$							L
	$P_u^{click\_15d}$						H	
	$P_u^{purchase\_15d}$		L	L		L		L
$TaxoDist_{rec}$	P_u^{click}						L	
	$P_u^{purchase}$			L	L	L		L
$TaxoDist_{avg}$	P_u^{click}							H
	$P_u^{purchase}$			H				H
$TaxoDist_{min}$	$P_u^{click\_15d}$			L		L		L

Note: 1). "Q', '"O", "C", "E", "A", and "N" respectively denote Curiosity, Openness to experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. 2). For significant interaction effects, "H" or "L" is given, meaning that in the High or Low group w.r.t. a specific user characteristic, the corresponding item feature has a greater impact on users' perceived serendipity.

5 DISCUSSION

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

RQ1: What item features can influence users' perceived serendipity?

Results about numerical item features show that lower item popularity, smaller time difference (the time distance between the current recommendation and the latest item of the same category in the user's profile), smaller category difference (the category level over the taxonomy that two items differ from), and shorter taxonomic distance (the minimum number of hops between two items' leaf categories over the taxonomy) (except the *average* variation) can be indicators of users' perceived higher serendipity of the recommendation. These findings, especially those involving item categories, seem to contradict our common assumption that a serendipitous recommendation should be from categories the user was not recently exposed to. In other words, it implies that items from (or similar to) categories the user has recently visited might still be considered unexpected in addition to being relevant. Moreover, we find that a short-term user profile (say 15 days) acts more effectively than a long-term profile (3 months) (except the average taxonomic distance), which may be because the user's tastes often change over time and hence her/his recent preference may more matter.

As for categorical item features, several high-confident association rules are identified, respectively regarding the top-category, the 2nd-level category, the leaf category, and the product type. Though the rules may sound trivial in some sense, they indicate that too much relevance among certain categories may lead users to perceive the recommendation less serendipitous. For instance, users, who ever purchased products from "Snacks / Nuts / Specialty" and "Makeup / Perfume / Beauty Tools" categories, tend to perceive a recommendation from " Women's clothing" of low serendipity.

RQ2: What user characteristics can affect their perceived serendipity, and furthermore how they would interact with item features to take the effect?

It shows that both *age* and *gender* can affect users' serendipity perception: Adult/older and/or male users are more likely to feel a recommendation serendipitous, while young females are harder to be surprised which we think may be because they are more familiar with the online shopping environment due to frequent browsing/purchasing experience. In addition, our results verify that users with high curiosity, openness to experience, conscientiousness, and/or extraversion are more likely to give a higher score on the recommendation's serendipity. More notably, we find that the other two personality traits also take significant effect: People with high neuroticism and/or low agreeableness are more inclined to possess higher serendipity perception. We believe the reason behind is that these five personality traits are all related to curiosity from different angles [14]: Openness to experience indicates a person's preference for variety, conscientiousness indicates her/his aim for achievement, extraversion indicates interest in the external object, agreeableness indicates altruism, and neuroticism indicates vulnerability.

In terms of the interaction effects between item features and user characteristics, there are two major observations: 1). Five item features are not affected by user characteristics, i.e., click profile based time difference, long-term click profile based minimal category difference, long-term click profile based minimal taxonomic distance, and short-term purchase profile based minimal taxonomic distance. 2). The other item features interact with different user characteristics. In particular, younger users are more sensitive to item popularity, while users with low curiosity, openness to experience, extraversion, or neuroticism are more sensitive to long-term purchase profile based recent category difference and short-term purchase profile based minimal category difference. Different interaction effects may reflect users' common behavior as influenced by their characteristics. For instance, because highly curious users would have often explored more diverse items, they are less sensitive to category difference than lowly curious users.

RQ3: What inspirations can this work bring to the design of serendipity-oriented recommender systems?

There are three design implications for the development of serendipity-oriented recommender systems, as inspired by the above findings: 1). In addition to unpopularity and item dissimilarity as emphasized in related work (see Section 2), we have particularly studied the effects of temporal feature (time difference) and category distance (category difference and taxonomic distance) as derived from timestamp information and taxonomy structure. We believe that those features could be helpfully integrated into the current recommendation algorithm to accommodate their roles in affecting users' serendipity perception. For instance, the cosine similarity function used in the traditional item-based CF algorithm might be replaced by the negative logarithmic function of category difference. Moreover, temporal feature and category distance could be possibly integrated to fulfill their combined effect. 2). To make the recommendation's serendipity degree more personalized, we may further consider users' personal characteristics, especially curiosity and personality traits. For example, for highly curious people, the "surprise" level of a recommendation might be strengthened to meet their propensity towards more unexpected discovery. 3). The significant interaction effects might be useful for adjusting the role of a certain item feature in consideration of one user characteristic. For example, still with the similarity function of CF as an example, we can fuse it with a particular personality trait, say neuroticism (as low neuroticism users are less sensitive to the change in category difference, a lower weight might be applied to the category difference with regard to this trait).

In future work, it should be worth verifying the actual performance of these implications in real systems.

6 LIMITATIONS

Our work has several limitations: 1). We focus on single-item serendipity in this work (each participant gave feedback on a single recommendation in our survey), but in reality, a user may receive multiple recommendations at one time. It should hence be interesting to assess users' perception of an entire list's overall serendipity as well. 2). The considered item taxonomy structure is mainly from the providers' perspective. Though we have attempted to include other product classification schemes that take into account consumers' evaluation approaches and involvement efforts, more studies should still be needed to identify the inherent relationship among categories within one structure, as well as the linkage among different structures from end-users' point of view. 3). More item features could be investigated, such as the combination of temporal feature and category distance as indicated above and the user's interaction frequency with a category. 4). Our studied domain is limited to e-commerce, so the generalizability of those findings to other domains should be validated in future work, especially those of different taxonomy structure and classification approaches. Last but not least, as discussed above, we will be interested in implementing the derived design implications in a concrete recommender system and measuring their practical effects on increasing the system's serendipity in comparison with the state-of-the-art.

REFERENCES

[1] Daniel Ellis Berlyne. 1960. *Conflict, Arousal and Curiosity*. McGraw-Hill.

[2] Toine Bogers and Lennart Björneborn. 2013. Micro-serendipity: Meaningful Coincidences in Everyday Life Shared on Twitter. In *Proceedings of the iConference 2013*. 196–208.

[3] Pablo Castells, Saúl Vargas, and Jun Wang. 2011. Novelty and Diversity Metrics for Recommender Systems: Choice, Discovery and Relevance. (2011), 8.

[4] Li Chen, Yonghua Yang, Ningxia Wang, Keping Yang, and Quan Yuan. 2019. How Serendipity Improves User Satisfaction with Recommendations? A Large-Scale User Evaluation. In *The World Wide Web Conference (WWW '19)*. ACM, New York, NY, USA, 240–250. https://doi.org/10.1145/3308558.3313469

[5] Yu-Shian Chiu, Kuei-Hong Lin, and Jia-Sin Chen. 2011. A Social Network-based Serendipity Recommender System. In *2011 International Symposium on Intelligent Signal Processing and Communications Systems (ISPACS)*. IEEE, 1–5.

[6] Paul T Costa and Robert R McCrae. 1992. NEO PI-R Professional Manual. Odessa, FL: Psychological Assessment Resources.

[7] Marco De Gemmis, Pasquale Lops, Giovanni Semeraro, and Cataldo Musto. 2015. An Investigation on the Serendipity Problem in Recommender Systems. *Information Processing & Management* 51, 5 (2015), 695–717.

[8] Patrick E. Murphy and Ben M. Enis. 1986. Classifying Products Strategically. Vol. 50, 3 (1986), 24–42. https://www.jstor.org/stable/1251583

[9] Gary T. Ford, Darlene B. Smith, and John L. Swasy. 1988. An Empirical Test of the Search, Experience and Credence Attributes Framework. NA-15 (1988). http://acrwebsite.org/volumes/6817/volumes/v15/NA-15

[10] Samuel D Gosling, Peter J Rentfrow, and William B Swann. 2003. A Very Brief Measure of the Big-Five Personality Domains. 37, 6 (2003), 504–528. https://doi.org/10.1016/S0092-6566(03)00046-1

[11] Dawn Iacobucci, Steven S. Posavac, Frank R. Kardes, Matthew J. Schneider, and Deidre L. Popovich. 2015. The Median Split: Robust, Refined, and Revived. *Journal of Consumer Psychology* 25, 4 (2015), 690–704. https://doi.org/10.1016/j.jcps.2015.06.014

[12] Carl Gustav Jung. 1983. *Memories, Dreams, Reflections*. London: Fontana Press. 448 pages.

[13] Marius Kaminskas and Derek Bridge. 2016. Diversity, Serendipity, Novelty, and Coverage: A Survey and Empirical Analysis of Beyond-Accuracy Objectives in Recommender Systems. *ACM Trans. Interact. Intell. Syst.* 7, 1, Article 2 (Dec. 2016), 42 pages. https://doi.org/10.1145/2926720

[14] Todd B. Kashdan, Matthew W. Gallagher, Paul J. Silvia, Beate P. Winterstein, William E. Breen, Daniel Terhar, and Michael F. Steger. 2009. The Curiosity and Exploration Inventory-II: Development, Factor Structure, and Psychometrics. *Journal of Research in Personality* 43, 6 (2009), 987–998. https://doi.org/10.1016/j.jrp.2009.04.011

[15] Noriaki Kawamae. 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, 218. https://doi.org/10.1145/1835449.1835487

[16] Noriaki Kawamae, Hitoshi Sakano, and Takeshi Yamada. 2009. Personalized Recommendation Based on the Personal Innovator Degree. In *Proceedings of the Third ACM Conference on Recommender Systems - RecSys '09*. ACM Press, 329–332. https://doi.org/10.1145/1639714.1639780

[17] Denis Kotkov, Joseph A. Konstan, Qian Zhao, and Jari Veijalainen. 2018. Investigating Serendipity in Recommender Systems Based on Real User Feedback. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing (SAC '18)*. ACM, New York, NY, USA, 1341–1350. https://doi.org/10.1145/3167132.3167276

[18] Denis Kotkov, Jari Veijalainen, and Shuaiqiang Wang. 2018. How Does Serendipity Affect Diversity in Recommender Systems? A Serendipity-Oriented Greedy Algorithm. (2018), 1–19. https://doi.org/10.1007/s00607-018-0687-5

[19] Denis Kotkov, Shuaiqiang Wang, and Jari Veijalainen. 2016. A Survey of Serendipity in Recommender Systems. *Knowledge-Based Systems* 111 (Nov. 2016), 180–192. https://doi.org/10.1016/j.knosys.2016.08.014

[20] Qiuxia Lu, Tianqi Chen, Weinan Zhang, Diyi Yang, and Yong Yu. 2012. Serendipitous Personalized Ranking for Top-N Recommendation. In *2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, Vol. 1. IEEE, 258–265.

[21] Valentina Maccatrozzo, Manon Terstall, Lora Aroyo, and Guus Schreiber. 2017. SIRUP: Serendipity In Recommendations via User Perceptions. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces (IUI '17)*. ACM, New York, NY, USA, 35–44. https://doi.org/10.1145/3025171.3025185

[22] Christian Matt, Thomas Hess, Alexander Benlian, and Christian Weiß. 2014. Escaping from the Filter Bubble? The Effects of Novelty and Serendipity on Users' Evaluations of Online Recommendations. (2014). https://EconPapers.repec.org/RePEc:dar:wpaper:66193

[23] Sean M. McNee, John Riedl, and Joseph A. Konstan. 2006. Being Accurate is Not Enough: How Accuracy Metrics Have Hurt Recommender Systems. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems (CHI EA '06)*. ACM, New York, NY, USA, 1097–1101. https://doi.org/10.1145/1125451.1125659

[24] Alan Menk, Laura Sebastia, and Rebeca Ferreira. 2017. CURUMIM: A Serendipitous Recommender System for Tourism Based on Human Curiosity. In *2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI)*. 788–795. https://doi.org/10.1109/ICTAI.2017.00124

[25] Makoto Nakatsuji, Yasuhiro Fujiwara, Akimichi Tanaka, Toshio Uchiyama, Ko Fujimura, and Toru Ishida. 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 (CIKM '10)*. ACM, New York, NY, USA, 949–958. https://doi.org/10.1145/1871437.1871558

[26] Phillip Nelson. 1970-03. Information and Consumer Behavior. 78, 2 (1970-03), 311–329. https://doi.org/10.1086/259630

[27] Xi Niu and Fakhri Abbas. 2017. A Framework for Computational Serendipity. In *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization - UMAP '17*. ACM Press, 360–363. https://doi.org/10.1145/3099023.3099097

[28] Kensuke Onuma, Hanghang Tong, and Christos Faloutsos. 2009. TANGENT: A Novel,'Surprise-me', Recommendation Algorithm. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 657–666.

[29] Gaurav Pandey, Denis Kotkov, and Alexander Semenov. 2018. Recommending Serendipitous Items using Transfer Learning. In *Proceedings of the 27th ACM international conference on information and knowledge management*. ACM, 1771–1774.

[30] Horace Walpole. 1960. To Mann, Monday 18 January 1754. In *Horace Walpole's Correspondence*. Yale University Press, 407–411.

[31] Chang-Dong Wang, Zhi-Hong Deng, Jian-Huang Lai, and Philip S. Yu. 2018. Serendipitous Recommendation in E-Commerce Using Innovator-Based Collaborative Filtering. *IEEE Transactions on Cybernetics* 49, 7 (2018), 2678–2692. https://doi.org/10.1109/TCYB.2018.2841924

[32] Yuan Cao Zhang, Diarmuid Ó Séaghdha, Daniele Quercia, and Tamas Jambor. 2012. Auralist: Introducing Serendipity into Music Recommendation. In *Proceedings of the 5th ACM International Conference on Web Search and Data Mining (WSDM '12)*. ACM, New York, NY, USA, 13–22. https://doi.org/10.1145/2124295.2124300

[33] Pengfei Zhao and Dik Lun Lee. 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, 315–324. https://doi.org/10.1145/2911451.2911488

[34] Qianru Zheng, Chi-Kong Chan, and Horace H. S. Ip. 2015. An Unexpectedness-Augmented Utility Model for Making Serendipitous Recommendation. In *Advances in Data Mining: Applications and Theoretical Aspects*, Petra Perner (Ed.). Springer International Publishing, Cham, 216–230.