Adaptive real-time diversification of digital content

Tushar Shandhilya
stushar@cse.iitk.ac.in
IIT Kanpur, India

Nisheeth Srivastava
nsrivast@cse.iitk.ac.in
IIT Kanpur, India

ABSTRACT

Diversifying recommendations is a vital component of maintaining high satisfaction with digital media. However, recommendation diversity must be sensitive to the context of a user's content browsing history. Someone who is exploring content narrowly focused on a specific area is unlikely to find tangential recommendations useful. Someone who has recently switched from browsing content of a particular category into another one is more likely to be interested in content of the second category. Finally, someone who is bored and browsing content randomly may prefer content with high intrinsic novelty. In this paper, we present an algorithmic technique for estimating peoples' novelty preference from browsing history, and demonstrate one possible way of adaptively introducing novelty which improves user experience. We conduct a blinded comparison study for this system vis-a-vis YouTube recommendations and show that participants prefer this system's recommendations in settings where browsing history is not focused on any specific category.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**; *User studies*; *Web-based interaction.*

KEYWORDS

datasets, Recommendations, Human factor, diversity

ACM Reference Format:
Tushar Shandhilya and Nisheeth Srivastava. 2022. Adaptive real-time diversification of digital content. In *Proceedings of The 2nd International Workshop on Online And Adaptive Recommender Systems with ACM SIGKDD Conference on Knowledge Discovery And Data Mining (OARS-KDD'22).* ACM, New York, NY, USA, 9 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Marshall McLuhan taught us that the medium is the message [18]. With a large fraction of the world's population now spending hours every day scrolling personalized media feeds for content [11], what exactly are the messages, embedded in the nature of the medium, they are receiving? As we are discovering to our discomfort, one of the meta-messages built into the nature of personalized digital feeds is the reinforcement of existing preferences [20] and reduction of exposure to sources of novel information [3].

OARS-KDD'22, Aug 14–18, 2022, Washington, DC, USA

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

Since digital media platforms must retain users' time on their site for as long as possible to maximize revenue from them, it is inevitable that UI design decisions, including the nature of content recommendations, are by and large be made in order to optimize this objective [25]. Consequently, content-based recommender systems, e.g. Netflix, inherently suffer from over-specialization - a tendency to present only the most typical, relevant and/or popular suggestions in response to user behavior [16]. Likewise, personalized recommender systems, e.g. YouTube, end up simply recommending content that the user has watched before [14]. Broadly, the design of recommender systems has found it difficult to escape from the behaviorist premise that people want to do what they have been seen to be doing before [6].

In ongoing efforts to engage with such limitations, the recommender systems (RS) research community has sought to algorithmically introduce diversity in recommendations to increase instances of serendipity in users' experience [15]. However, most of these approaches have focused on methods of introducing diversity as an explicit quota system, wherein content rankings continue to primarily be generated using users' previously revealed preferences, but are augmented with a small sprinkling of novel content from lower down in the ranking [23]. Similarly, [7] introduce diversity by prioritizing long-tail items in recommendations to support their primary focus on relevance.

Comparatively less attention has been paid to the psychological dynamics [5, 9] underpinning the pattern of preferences shown by human content consumers [4]. Cognitive psychology research findings suggest that this pattern tends to follow sequential stages of introduction of new content, appreciation or rejection, sustained engagement for a period of time, followed by disinterest, followed in turn by a 'fallow' period after which interest in the content may be rekindled (or not) [1, 2].

Lack of constructive engagement with theoretically well-understood dynamics of human preferences has considerably hindered efforts at introducing novelty in peoples' recommendation streams *adaptively* [17]. Intuitively, it is evident that someone deeply engaged with sampling content in a particular category would prefer recommendations specific to that category. On the other hand, someone exploring content without strong prior preferences may prefer greater diversity in their recommendation lists. Thus, users' desire for novelty in recommendations is both context- and time-sensitive, suggesting that diversification of recommendations must be adaptive to such dynamic changes in users' novelty preference. In this paper, we show how to estimate users' preference for novelty based on their past behavior, and how to deliver novel content adaptively using such estimates, thus offering adaptive diversification of recommendations.

The remainder of this paper is structured as follows: In Section 2, we describe our theoretical and modeling approach to solving the adaptive diversification problem. In Section 3, we describe a novel dataset we collected, that contains all desirable data features necessary for us to train our model's parameters. In Section 4, we describe the statistical calculations needed to fit the model's parameters to data. Section 5 demonstrates the ability of our model, trained on our dataset, to adaptively tune the diversity of it's recommendations based on a user's browsing history. Section 6 reports results from a blinded comparison study we conducted a on a small panel of internet users, and Section 7 concludes with an appraisal of the overall contribution of this paper.

2 MODELLING

Following the framework developed in [24], we can formalize the problem of providing adaptive recommendations as follows. Consider the recommendation list R_u expressed in terms of Similarity list S_u and Novelty list N_u for an user u where λ is the trade-off between similar and novel recommendations.

$$R_u = \frac{S_u + \lambda * N_u}{1 + \lambda} \quad (1)$$

The problem of making adaptive recommendations can be understood then as the task of estimating λ. In classic diversification approaches, like [24], λ is considered to be constant. Recently, [12] introduced a regression model to predict users' λ using past behavior. We build upon the approach proposed in [12] to design a method for adaptively diversifying recommendations.

2.1 Modelling user histories with HSMMs

Kapoor et al. [13] have empirically demonstrated that boredom significantly affects the temporal dynamics of consumption of familiar songs on radio, such that users shift to different songs when bored with it after multiple auditions and come back to it later with interest restored by time spent not listening to it [13], consistent with the theoretical expectations of classic accounts of novelty preference [2]. They also proposed that this pattern of behavior can be modelled using a hidden semi-Markov model (HSMM) with two psychological latent states of sensitization and boredom. Users transition from sensitization to boredom with multiple frequent exposures to the same song, and from boredom to sensitivity given sufficient time away from it [13].

Our modelling builds on the same premise: we model user behavior while sequentially browsing web content using a HSMM, with the user switching between latent psychological states of *sensitization* to a particular item (or item) and *boredom* with it, as illustrated in Figure 1. Estimating the transition probabilities and durations of this process for each user based on their existing browsing history gives us a way of predicting their preferences for novel content in future. We describe below how we do this in more detail.

Following [12], we assume that when the user is sensitive to an item they consume the item more frequently in regular intervals whereas when the user finds boredom in the item the visits take more time which later leads to forgetting the item.

As shown in figure 2, let g_x^{ui} be the gap (time interval) after which a user visits an item i sampled from a random variable G, and D be the random variable for duration a user stays in either state (B , S) for a particular item. Using {G, D} estimated from browsing histories, we describe below how an HSMM model can predict future novelty preference.

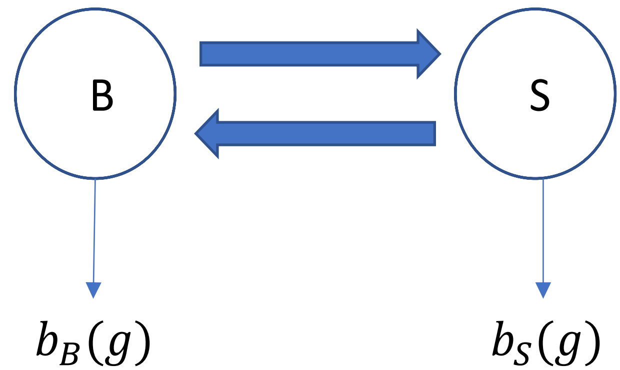

Figure 1: An HSMM model of content browsing, with latent variables representing the psychological states of boredom(B) and sensitization(S) with/to content, with corresponding emission densities.

Formally, in an HSMM, a state is associated with an emission density $b_m(g) = P(G = g|m)$ for $m \in \{B, S\}$ where $P(.)$ is a state-conditioned distribution on the gap-length random variable (G) and $B_m(g)$ is cumulative distribution. The likelihood of the observed output is $P(\{g, \delta\}|m) = (1 - \delta) * b_m(g) + \delta * (1 - B_m(g))$ where δ is a special status variable which is use to censor time gaps larger than a threshold value (indicating session breaks) in the data.

The parametric form of the state distributions and emission are $p_m(d) = Gamma(\alpha_m, \beta_m)$ and $b_m(log(g)) = Log\text{–}logistic(\mu_m, \sigma_m)$ respectively. The parameters of the model are given by Λ = (A, π, $b_m(g)$, $p_m(d)$), where A is the transition probability over m states, π denotes the initial state probability distribution of m states.

The distribution of the next latent state is given by $s_n(m) = P(s_n = m|g_{1...(n-1),\Lambda})$, $s_n(S) = 1 - s_n(B)$ because the model has only two latent states.

Having concretely operationalized HSMM model parameters using gap and visitation information accessible from users' browsing history of participants, these parameters (Λ) can now be estimated through maximum likelihood estimation using the forward-backward algorithm for all participants [23].

2.2 Quantifying diversity and novelty

Several mutually consistent methods have been proposed to measure users' experienced diversity in the system [15]. We define diversity div_t^u for a user u at time t as the number of categories of items seen by the user in a browsing session divided by the number of items viewed in the session, a very common operationalization of diversity in the literature [15].

While novelty can be naively instantiated as an inverse of frequency of item exposure, such a definition misses important psychological desiderata of novelty expectations [2]. From a psychological perspective, the triggering of conflict of experience with existing mental schemata is a defining property of the experience of novelty [1]. Infrequency of exposure is a necessary, but not sufficient condition for the generation of such conflict; content that conforms

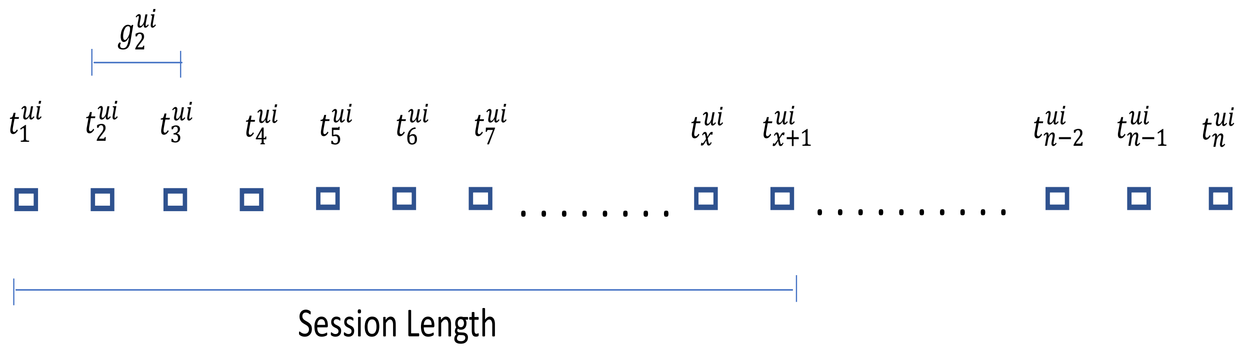

Figure 2: This figure shows an abstract representation of the browsing history of a user u of an item i at different times t, the gap between two timestamps g_x^{ui} and an individual session's boundaries.

to users' prior expectations will not register psychologically as novel, even if it has not actually been previously seen.

Recently [19] introduced incongruity as a method for ranking items likely to increase the diversity of users' recommendation experiences. They operationalize incongruity as the probability of violating contextual expectations, for example, the concept 'Messi buys a football in a sports shop' is normal, whereas 'Messi buys a pencil in a sports shop' is incongruous, because the presence of 'pencil' is incongruous in the context of purchases from sports shop. They measure incongruity using semantic word embeddings, and show that using this measure caused users to accept significantly more recommendations while exploring Wikipedia.

$$\begin{array}{r c} & \begin{array}{cccccc} \text{Cat} & \text{Milk} & \text{Mouse} & \text{Chase} & \text{nip} & \text{Champagane} \end{array} \\ \begin{array}{r} \text{Cat} \\ \text{Milk} \\ \text{Mouse} \\ \text{Chase} \\ \text{nip} \\ \text{Champagane} \end{array} & \begin{bmatrix} 1 & 21 & 13 & 30 & 33 & 89 \\ 21 & 1 & 26 & 39 & 37 & 75 \\ 13 & 26 & 1 & 22 & 29 & 80 \\ 30 & 39 & 22 & 1 & 15 & 91 \\ 33 & 37 & 29 & 15 & 1 & 87 \\ 89 & 75 & 80 & 91 & 87 & 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 186 \\ 198 \\ 170 \\ 197 \\ 201 \\ \color{red}{422} \end{bmatrix} \end{array}$$

Figure 3: This figure shows a sample example of incongruity score calculation.

Given its alignment with theoretical desiderata for the experience of novelty, in this work, we use incongruity as our mechanism for introducing novelty in recommendations. As shown in Figure 3 the incongruity score is calculated by taking keywords associated with the item and finding a distance matrix which contains pairwise semantic distances. From the distance matrix, we calculate a row-wise average, ignoring the self distance. From the vector obtained, we find the number of words which are k standard deviations away from the average median of the vector, which are considered anomalous to the context. As illustrated in Figure 3, the term 'champagne' is anomalous in a context otherwise suggested by the set of keywords 'cat , 'milk', 'mouse', 'chase' and 'nip'. The number of anomalous words divided by the total number of words gives the incongruity score. The choice of k is a domain-specific free parameter; we used $k = 2$ for our experiments. We incorporate this concept within our approach to ensure that items with high novelty are presented to users when our system estimates their novelty-preference to be high.

2.3 Predicting user preferences

Sensitization and boredom from the items drive the novelty needs of the user, the novelty needs dictate the preference of the item by the user. Using HSMM model defined in section 2.1 we model this user behaviour, the model can be used to quantify the dynamic preference of an item as a function of time [13].

The dynamic preference score ($DPref_i(t)$) gives the likelihood of the user consuming the item at time t, given the model parameters and gaps sequence which is denoted by . We find the Dynamic preference score for the time t as follows:

$$DPref_i(t) = \frac{s_n(S) * b_S(t - t_n) + s_n(B) * b_B(t - t_n)}{s_n(S) * (1 - B_S(t - t_n)) + s_n(B) * (1 - B_B(t - t_n))}. \tag{2}$$

Following [12] , in a session $[t - T, t)$ the user views many items. The overall boredom of the user is modelled as the cumulative negative transform of the dynamic preference score for each item in the familiar set of the user at time t,

$$NCP_t^u = \sum_{i \in famset_u^t} -DPref_t^{(i,u)}, \tag{3}$$

reflecting the intuition that someone with a large negative preference for familiar items is bored. Thus, NCP_t^u directly quantifies boredom at time t for user u in our model.

3 THE YOUTUBE PREFERENCE DYNAMICS DATASET

In seeking to accommodate psychological aspects of preference dynamics into the estimation of novelty preference and satiation, we have presented a modelling approach that demands multiple forms of data. It requires data about the time and duration for which people engage with content, over long periods of times and across multiple sessions. In order to make psychologically meaningful predictions, content items must belong to distinctive conceptual categories, such that the latent variables of content-sensitivity and

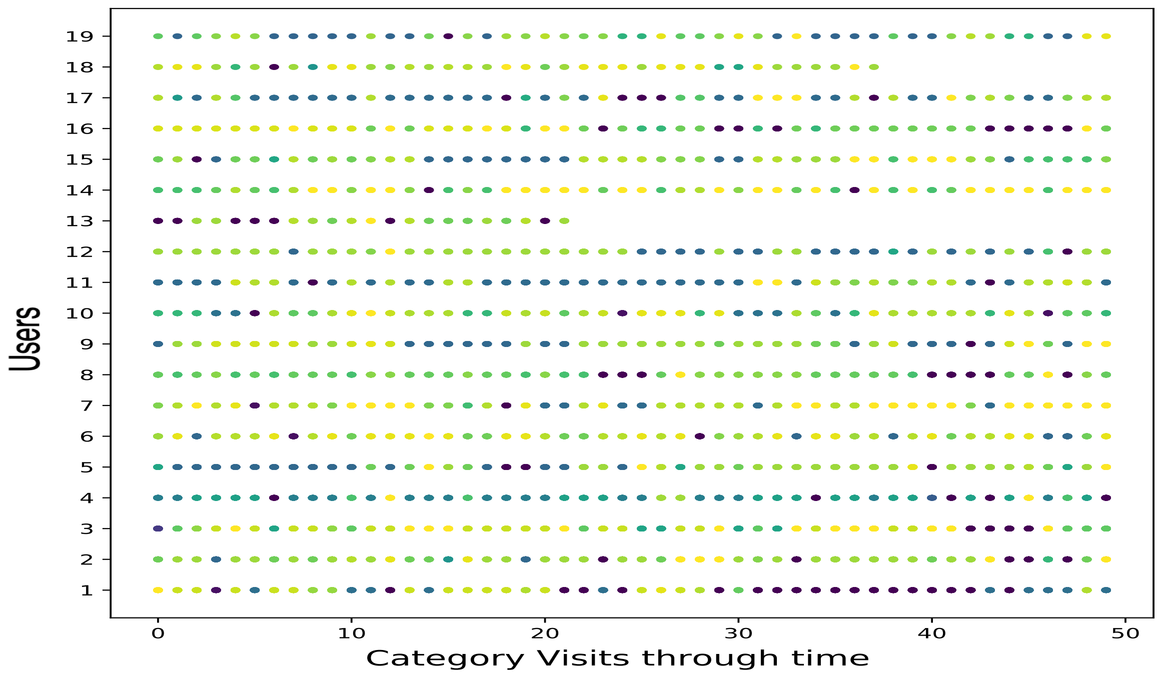

Figure 4: This figure shows typical video browsing behavior of 20 randomly selected users in our dataset, abstracted out at the level of visitation of videos belonging to specific categories by users with respect to time, Colors of the dots represent different categories visited by the user. Notice the blocky nature of most users' browsing behavior.

boredom may be plausibly operationalized. Additionally, these content items cannot be semantically atomic, they must have semantic associations that can be used to calculate incongruity scores.

To satisfy these information requirements, we have collected a novel dataset, consisting of watch histories of YouTube users. YouTube, and other similar streaming media platforms, provide content curated by conceptual categories, and alongside platform- or user-generated meta-data that can be mined for semantic associations.

We used the following data collection procedure. We collected watch histories from about 50 different users (21F, Median age = 23 years) with equal number of males and females. The users were asked to share their YouTube history of past 2 years which the user can access from their google takeout account. They sent us their watch-history HTML files. We extracted links of the videos, date and time of when they were watched using headless browsing scripts. Using the links of the video and the YouTube API we obtained the category for each video. A sample snapshot of data from an user is shown below in Table 1

In our dataset each user has consumed 4300 videos on an average and as can be seen in Fig. 4 the users ordinarily visit 3 - 4 categories very frequently and on an average, visited about 11 categories across their entire watch histories (out of 32 total categories).

Youtube video link	Timestamp	Category
Video Link 1	1610352394.0	25
Video Link 2	1610352459.0	24
Video Link 3	1610355332.0	26
Video Link 4	1610355404.0	27
Video Link 5	1610355441.0	26
Video Link 6	1610378542.0	28
Video Link 7	1610378744.0	22
Video Link 8	1610420624.0	23
Video Link 9	1610425566.0	24
Video Link 10	1610425598.0	24

Table 1: This figure shows the snapshot of a users' watching history containing youtube video link, watching timestamp, Category id which is extracted for a youtube link using YouTube API

We also observe that, on average, users demonstrate non-random blocky behavior, wherein they watch 4-6 videos of the same category consecutively, after which the user either switches to a new category or keep switching between different categories, consistent with our theoretical expectations as well as with the use of HSMMs as a model of user browsing behavior.

In the anticipation that this rich and large dataset will be useful in studies beyond ours, we are publicly releasing it on OSF. The data can be found here.

4 MODEL ESTIMATION

In this section, we describe how we fit our model specifically to the YouTube Preference Dynamics dataset we describe above.

4.1 HSMM model fitting

Our dataset contains the category of each video, which we consider as an item i an user visits for our purposes. Treating each video as an individual item would be unrealistic in this case, since most videos are viewed only once, and very few videos are viewed by multiple users. Fitting the model at the level of topic categories permits more generalizable conclusions to be extracted, both within and across users. As shown in fig. 2 the visiting history of the user for an item i is $H^{ui} = \{t_1^{ui}, t_2^{ui}, t_3^{ui}, t_4^{ui}, t_5^{ui}, t_6^{ui}, ..., t_n^{u}\}$, where t_n^{ui} is the last consumption of the item by the user . The gaps g_x^{ui} after which an user u visits an item i is $t_{x+1}^u - t_x^u$ where $x \in \{1, 2, 3, ..., n-1\}$ and $g_n^{ui} = \{T - t_n^{ui}\}$ where observation time for a session ends at time T .

At any time t the session (S_t^u) is the continuous period in which the user u consumes different items with a small time gap , the total number of possible items is I.

We censored gaps which are greater than a threshold value of 15 minutes. A status variable δ_t^{ui} is used to handle the censored gaps where 0 is assigned to the censored gap and 1 otherwise.

After this pre-processing, HSMM model parameters(Λ) were estimated through maximum likelihood estimation using the forward-backward algorithm for all participants [23].

4.2 Measuring diversity

Let $famSet_t^u$ be the set of categories visited by the user in the observation time T, categories consumed less than a threshold are removed to eliminate noise. According to [13] the diversity of the consumed items (familiar items) by the user is given by: $div_t^u = \frac{A}{B}$ where A = $|famSet_t^u|$ and B = Number of times items of $famSet_t^u$ were viewed in $[\, t - T, t]$. The novel set $(nvSet_t^u)$ consumed by the user is identified as $I - famSet_t^u$.

4.3 Calculating incongruity of videos

We use meta-data (tags) associated with YouTube videos obtained using YouTube API as mentioned in section 3 to calculate the incongruity scores for each YouTube video seen by each user in our dataset, following the procedure mentioned in Section 2.2. ICG_t^u is the average incongruity score of all the videos watched by a user in the observation period $[t - T, t)$.

4.4 Novelty preference estimation

In an observation period T the ground novel preference by the user u at a time t is given by:

$$nvPref_t^u = \frac{|S_t^u \cap nvSet_t^u| + c}{|S_t^u| + 2 * c} \tag{4}$$

Case No	$nvPref$
5.1	0.324
5.2	0.561
5.3	0.801

Table 2: This table gives the $nvPref$ scores obtained for different case studies mentioned in 5

where c is a Laplacian correction for small sessions.
We train a logistic regression model to calculate the predicted $nvPref$ score.

$$nvPref = logistic(div_t^u, NCP_t^u, ICG_t^u) \tag{5}$$

where $NCP_t^u = \sum_{i \in famSet_t^u} -DPref_t^{(i,u)}$ and ICG_t^u is the average Incongruity score of all the videos watched in the observation period as described in section 2.3 and section 4.3 respectively.

Finally, we use the $nvPref$ scores calculated in 5 to modulate λ in 1 to rank recommendations for the user.

The model estimates novelty preference partly as a function of observed incongruity preference, stimulated by the close correlation between novelty and incongruity demonstrated recently in [19]. It privileges high incongruity items to display at times of high novelty preference for users with a history of preferring incongruous items.

4.5 Making recommendations

Obtaining the $nvPref_t^u$ scores using the logistic regression models, the recommendations at some time t can be given using :

$$R_{u,t} = \frac{S_{u,t} + \lambda * N_{u,t}}{1 + \lambda} \tag{6}$$

where λ is identical to $nvPref_t^u$.

For our case studies and user study, we selected a corpus of about 10000 videos, evenly sampled from all YouTube categories. From this corpus, we obtain $S_{u,t}$ as YouTube videos matched by similarity to the watch history of the user u in the time range $[t - T, t)$ from among the set of video categories that are members of $famSet_t^u$. Likewise, $N_{u,t}$ is obtained by ranking videos from $nvSet_t^u$ categories using incongruity score [19] calculated using meta-data (tags) of the videos, where $nvSet_t^u$ is the set of video categories unseen by user u upto time t from the categories represented in the corpus. Recommendations from within $N_{u,t}$ are made using incongruity as the ranking criterion, and from within $S_{u,t}$ by random sampling.

5 CASE STUDIES

Applying the trained model parameters to individual users' browsing histories reveals interesting insights. We demonstrate three representative case studies using small segments of watch histories randomly extracted from our dataset and covering important use cases below, showing the algorithm's ability to correctly estimate novelty preference, and offer suitably adapted recommendations. All case examples were obtained using novelty preference estimates calculated using a user's browsing history for the particular session we show in the corresponding case example as 'History'. The actual novelty preference values estimated for the case examples are given in Table 2.

History	Introduction to For Loops in Python (Python Tutorial #5)
	How To Use Functions In Python (Python Tutorial #3)
	Introduction to Classes and Objects - Part 1 (Data Structures & Algorithms #3)
	While Loops and The Break Statement in Python (Python Tutorial #6)
	Introduction To Lists In Python (Python Tutorial #4)
	Making a Snake Game Where You're the Food in Python
	Introduction to Linked Lists (Data Structures & Algorithms #5)
	How To Use Sets in Python (Python Tutorial #13)
	List Comprehension Basics with Python (Python Tutorial #12)
Model Recommendations	Floating Point Numbers - Computerphile
	But what is a Neural Network? | Deep learning, chapter 1
	Whatsapp System Design: Chat Messaging Systems for Interviews
	Python Machine Learning Tutorial (Data Science)
	AlphaGo - The Movie | Full Documentary
	How to Learn to Code and Make $60k+ a Year
	Introduction to Linked Lists (Data Structures & Algorithms 5)
	How To Think Like A Programmer
	Python WiFi

Table 3: This figure shows watch history and recommendations given by our model for a user engaged in watching a single category of videos in a session.

5.1 Focused Browsing

The History row in Table 3 documents the recent browsing history of a user who evidently is engaged in learning to code in python. While browsing in a focused manner, users generally prefer few distractions in recommendations, whether produced overtly in response to a search query, or covertly in the sidebar as is seen in YouTube's design.

As we see in Table 2, our model correctly assesses this as a low novelty preference scenario, and produces recommendations that are mostly centered around programming videos. We note that the single incongruous suggestion made in the list, the AlphaGo documentary, is also about a programming achievement. These suggestions demonstrate both the value of the novelty preference in producing a relevant list of suggestions and the value of incongruity as a way of producing thematically related novel suggestions likely to create serendipity.

5.2 Transitional Browsing

A user watching videos (as shown in Fig. 4) related to programming (Category 27 Education) switches to watching Dance videos (Category 24 Entertainment).

The watch history documented in Table 4 presents an interesting example of a situation wherein the user may have switched away from a specific work-related focus (programming) to an equally specific leisure-related focus (dance) either by conscious choice or through a failure of willpower.

This case presents a clear demonstration of a limitation of our current approach. Whereas the watch history clearly indicates a switch from one focus to another, our model interprets the presence of two categories in the user's recent watch history as evidence for more novelty seeking than usual, and populates the recommendation list with more novel recommendations. However, we note that the user can also still see programming and dance related videos in the recommendation list since the novelty preference value is not exceptionally high.

5.3 Exploratory Browsing

In the watch history shown in Table 5, there is no clear focus seen on any single category, suggesting that the user is browsing in an exploratory manner. Our model estimates a high novelty preference for this session for this user and recommends a diverse set of videos as a consequence.

In general, we expect that the modes of behavior represented by the three case studies demonstrated above fairly comprehensively cover the set of browsing behaviors relevant for adaptive recommendations. People will either be focused narrowly on one category, be switching deliberately from one category to another, or browsing in an exploratory manner. Our model is sensitive to the differences between these categories, and adapts recommendations to these modes accordingly such that users who are in work-mode see only work-relevant recommendations, and users in play-mode see a more diverse set. However, our model fails to behave intelligently in situations wherein people are transitioning in a focused manner from one category to the next, suggesting an obvious scope for improvement.

Browsing the videos in focused manner in a particular area may lead to rabbit hole effect [22] where the user can keep watching the videos and get hooked. This behaviour can be detrimental or useful for the user, we believe this is a conscious decision an user should take while surfing the data so that it doesn't have long term adverse effect on the user.

6 USER STUDY

The case studies presented above demonstrate that our modelling approach is able to track users' novelty preference, and adjust recommendations accordingly. However, they do not establish that the recommendations made in this manner are likely to be more desirable for users than non-adaptive recommendations made by existing systems. To test whether this was true, we conducted a user study to see whether user gets interested more by YouTube's native recommendations or our system's recommendations for case examples representative of the three modes of browsing seen above. We conducted this study as a classic A/B test, pitting our

History	Minimum Window Substring - Airbnb Interview Question - Leetcode
	Course Schedule - Graph Adjacency List - Leetcode 207
	Find Median from Data Stream - Heap & Priority Queue - Leetcode
	Word Search - Backtracking - Leetcode 79 - Python
	Jump Game II - Greedy - Leetcode 45 - Python
	B-Boy Lil G's BEST moments | 10 YEARS of Red Bull BC One All Stars
	Royal Family | FRONTROW | World of Dance Los Angeles 2015 | #WODLA15
	Complot Imperium | Team Division | World of Dance Championships 2018 | #WODCHAMPS18
Model Recommendations	BEST Dance Group on America's Got Talent 2019? | Got Talent Global
	8 Mile | Eminem's Final Rap Battles
	Writing a Python Script to Control my Lights | Five Minute Python Scripts
	BEST ALL CHAIR TURN Blind Auditions in The Voice
	LeetCode Peak Index in a Mountain Array Explained - Java
	What REALLY is Data Science? Told by a Data Scientist
	JABBAWOCKEEZ at the NBA Finals 2019
	TOP 10 Best Powermoves Sets of 2018

Table 4: This figure shows watch history and recommendations given by our model for a user who switches from one category to another during a session.

History	15 Most Expensive Dogs in the World
	NEW WORLD RECORD in SEASON 17!! | 53 KILLS vs SQUADS | PUBG Mobile
	Krystal Ball: Bill Gates Is LYING TO YOU On Vaccine Patent Protection
	Man in court for Lens woman's murder
	RR vs SRH full highlight match
	1. Introduction to Human Behavioral Biology
	The 116 images NASA wants aliens to see
	The art of cognitive blindspots | Kyle Eschen | TEDxVienna
Model Recommendations	Top Gun: Maverick (2021) â“ New Trailer - Paramount Picture
	Here's how to destroy your marriage on Family Feud!
	Record Breakers! Chris Gayle & Evin Lewis chase down 129 runs in just 7 overs! | CPL 2017
	The 116 images NASA wants aliens to see
	Battleship | The Final Battle in 4K HDR
	1095. Find in Mountain Array (LeetCode Weekly Contest 142)
	COVID-19: 1 Month Still Infected, What Now?
	Horror Short Film "The Smiling Man" | ALTER

Table 5: This figure shows watch history and recommendations given by our model for a user browsing videos in an exploratory manner.

model's content recommendations against YouTube's live content recommendations.

6.1 Design

Each question in the user study consisted of three columns where the first column contained the history of videos (with thumbnails) watched by a user and the other two columns contained the recommendations obtained from YouTube's native engine and our engine, counter-balanced for position both within and between participants and with no identifying labels for the source of the recommendations. We selected 3 questions each similar to each of the three case studies 5.1, 5.2, 5.3 making it a study of 9 questions.To get YouTube's native recommendations as listed in the study we made a new YouTube account for each question, watched the videos presented in the watch history in sequence, and extracted screenshots of YouTube's recommendations at the end of watching all the videos for each case example, clearing the watch history of the user account after each case's recommendations had been acquired.

For each trial, participants were asked to look at the watch history shown in the left column and asked to consider it to have been their own recent watch history. Taking this perspective, they were asked to select the list of recommendations they thought was better from between the center and right columns. This setup considers the complexity of displaying recommendation to the user as youtube since it contains the thumbnails and titles similar to live youtube recommendations.

We measure preference for each recommender system by the number of times users selected it, combined across all questions and participants. A priori, we expected that users would find our recommender system more desirable when they were being exploratory, but would prefer the YouTube system, a very strong baseline contrasted with our simple similarity-based relevance recommender, for focused browsing.

6.2 Participants

Sixty convenience-sampled users participated in our user study where 40 percent identified their gender as females rest 60 percent

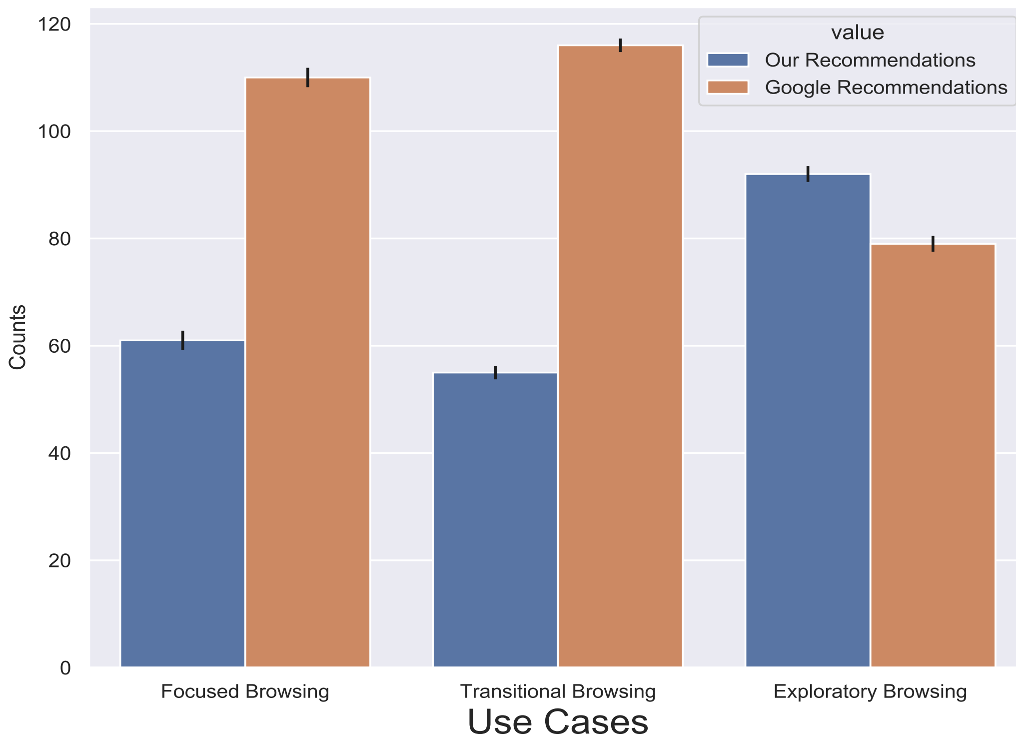

Figure 5: This figure shows the bar plot between the count of users vs the use cases of our model's recommendations and Google's recommendations. Error bars represent ± 1 SEM.

identified as males . The majority of participants were college students with an average age of 21. All data collection was approved by a university IRB.

6.3 Results

As shown in Fig. 5 we find that both in focused category-wise browsing (Cohen's d = 0.93) and for focused switching between categories (Cohen's d = 1.24), people in our panel strongly preferred YouTube's recommendations whereas for exploratory browsing, people preferred our system's recommendations (Cohen's d = 0.30). Thus, people behave precisely as expected in the study. They prefer the native YouTube engine for recommendations supporting focused browsing, but prefer our engine for recommendations during exploratory browsing. Thus, the outcomes of the study are entirely consistent with our expectations of user behavior, and support the value of our approach in offering novel recommendations when they're needed, and not otherwise.

An anonymized version of the user study itself, for reference, can be found here.

7 CONCLUSION

In this paper, we make three contributions. One, we develop a theoretical approach for adaptive diversification of online content by modeling the psychological dynamics of boredom and satiety for users browsing digital content. Building on previous work that documented how to estimate users' novelty preference from browsing histories, we developed an approach that offers users interested in novelty at any point in time recommended items that are intrinsically interesting by virtue of being both conceptually relevant with respect to the user's existing watch history and conceptually incongruous. Two, we present a new dataset of preference dynamics in video browsing history and methods for operationalizing our theoretical approach with this dataset to predict when users might be interested in novelty, and what to recommend to them based on their individual watch histories. Three, we present results from a user study documenting that, when people are browsing the web with no clear informational intent, recommendations built around a combination of topical relevance and conceptual incongruity are considered preferable to contemporaneous YouTube recommendations.

Several directions for improvement upon the work reported in this paper suggest themselves. Improving upon the construction of the latent variable model we have used to differentiate switches away from a category based on explicit choice versus switching away due to a lapse in executive control is one such direction. Assessing the efficacy of adaptive diversification in a live streaming setting is another, although this is very unlikely to be possible in academic settings.

As we discussed in the introduction of this paper, the recommender systems community is deeply interested in designing interfaces and algorithms that keep people from being trapped inside positive feedback loops [15, 16]. Researchers have had some success in designing spatial interfaces for promoting greater exploration of recommendations, such as [21] who show that a multidimensional visual interface promotes exploration of diverse recommendation. An excellent example of such an interface is OpinionSpace, a two-dimensional presentation of comments on web forums, wherein users can browse the space to select comments reflecting a diversity of opinions [8]. Such effort are clearly complementary to our work, in that they work on improving interfaces in space, while we do so in time.

Whereas some amelioration of the trap of behaviorism can be expected via thoughtful design of spatial interfaces [10, 21], information retrieval algorithms must also learn to engage with humans' preference dynamics in psychologically realistic ways [6]. Our work demonstrates a practical recommendation system that diversifies content in ways sensitive to psychological reality, as documented in earlier research [12, 19]. Adoption of this approach, or improvements thereupon, should help in the construction of recommender systems sensitive to the natural temporal rhythms of peoples' preferences.

REFERENCES

[1] Daniel E Berlyne. 1960. Conflict, arousal, and curiosity. (1960).

[2] Daniel E Berlyne. 1970. Novelty, complexity, and hedonic value. *Perception & psychophysics* 8, 5 (1970), 279–286.

[3] Homanga Bharadhwaj and Nisheeth Srivastava. 2019. New Tab Page Recommendations Strongly Concentrate Web Browsing to Familiar Sources. In *Proceedings of the 10th ACM Conference on Web Science*. 7–16.

[4] Pradeep K Chintagunta. 1998. Inertia and variety seeking in a model of brand-purchase timing. *Marketing Science* 17, 3 (1998), 253–270.

[5] William N Dember and Harry Fowler. 1958. Spontaneous alternation behavior. *Psychological bulletin* 55, 6 (1958), 412.

[6] Michael D Ekstrand and Martijn C Willemsen. 2016. Behaviorism is not enough: better recommendations through listening to users. In *Proceedings of the 10th ACM conference on recommender systems*. 221–224.

[7] Farzad Eskandanian and Bamshad Mobasher. 2020. Using Stable Matching to Optimize the Balance between Accuracy and Diversity in Recommendation. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization* (Genoa, Italy) *(UMAP '20)*. Association for Computing Machinery, New York, NY, USA, 71–79. https://doi.org/10.1145/3340631.3394858

[8] Siamak Faridani, Ephrat Bitton, Kimiko Ryokai, and Ken Goldberg. 2010. Opinion space: a scalable tool for browsing online comments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1175–1184.

[9] Murray Glanzer. 1953. The role of stimulus satiation in spontaneous alternation. *Journal of experimental psychology* 45, 6 (1953), 387.

[10] Marti A Hearst, David R Karger, and Jan O Pedersen. 1995. Scatter/gather as a tool for the navigation of retrieval results. In *Working Notes of the AAAI Fall Symposium on AI Applications in Knowledge Navigation and Retrieval, Cambridge, MA*.

[11] OECD Social Indicators. 2019. Society at a Glance 2019. *Recuperado de http://www. oecd. org/social/society-at-a-glance-19991290. htm* (2019).

[12] Komal Kapoor, Vikas Kumar, Loren Terveen, Joseph A. Konstan, and Paul Schrater. 2015. "I like to Explore Sometimes": Adapting to Dynamic User Novelty Preferences. In *Proceedings of the 9th ACM Conference on Recommender Systems* (Vienna, Austria) *(RecSys '15)*. Association for Computing Machinery, New York, NY, USA, 19–26. https://doi.org/10.1145/2792838.2800172

[13] Komal Kapoor, Karthik Subbian, Jaideep Srivastava, and Paul Schrater. 2015. Just in Time Recommendations: Modeling the Dynamics of Boredom in Activity Streams. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining* (Shanghai, China) *(WSDM '15)*. Association for Computing Machinery, New York, NY, USA, 233–242. https://doi.org/10.1145/2684822.2685306

[14] Denis Kotkov, Jari Veijalainen, and Shuaiqiang Wang. 2016. Challenges of serendipity in recommender systems. In *WEBIST 2016: Proceedings of the 12th International conference on web information systems and technologies. Volume 2, ISBN 978-989-758-186-1*. SCITEPRESS.

[15] Denis Kotkov, Shuaiqiang Wang, and Jari Veijalainen. 2016. A survey of serendipity in recommender systems. *Knowledge-Based Systems* 111 (2016), 180–192.

[16] Pasquale Lops, Marco De Gemmis, and Giovanni Semeraro. 2011. Content-based recommender systems: State of the art and trends. In *Recommender systems handbook*. Springer, 73–105.

[17] Kai Lukoff, Ulrik Lyngs, Himanshu Zade, J Vera Liao, James Choi, Kaiyue Fan, Sean A Munson, and Alexis Hiniker. 2021. How the design of youtube influences user sense of agency. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–17.

[18] Marshall McLuhan. 1994. Understanding Media: The Extensions of Man. (1994).

[19] Tushar Shandhilya and Nisheeth Srivastava. 2020. Using Conceptual Incongruity as a Basis for Making Recommendations. In *Fourteenth ACM Conference on Recommender Systems* (Virtual Event, Brazil) *(RecSys '20)*. Association for Computing Machinery, New York, NY, USA, 557–561. https://doi.org/10.1145/3383313.3412231

[20] Dominic Spohr. 2017. Fake news and ideological polarization: Filter bubbles and selective exposure on social media. *Business Information Review* 34, 3 (2017), 150–160.

[21] Chun-Hua Tsai and Peter Brusilovsky. 2017. Leveraging Interfaces to Improve Recommendation Diversity. In *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization* (Bratislava, Slovakia) *(UMAP '17)*. Association for Computing Machinery, New York, NY, USA, 65–70. https://doi.org/10.1145/3099023.3099073

[22] Kaitlin Woolley and Marissa A. Sharif. 2022. Down a Rabbit Hole: How Prior Media Consumption Shapes Subsequent Media Consumption. *Journal of Marketing Research* 59, 3 (2022), 453–471. https://doi.org/10.1177/00222437211055403 arXiv:https://doi.org/10.1177/00222437211055403

[23] Shun-Zheng Yu and H. Kobayashi. 2003. An efficient forward-backward algorithm for an explicit-duration hidden Markov model. *IEEE Signal Processing Letters* 10, 1 (2003), 11–14. https://doi.org/10.1109/LSP.2002.806705

[24] Tao Zhou, Zoltán Kuscsik, Jian-Guo Liu, Matúš Medo, Joseph Rushton Wakeling, and Yi-Cheng Zhang. 2010. Solving the apparent diversity-accuracy dilemma of recommender systems. *Proceedings of the National Academy of Sciences* 107, 10 (2010), 4511–4515.

[25] Lixin Zou, Long Xia, Zhuoye Ding, Jiaxing Song, Weidong Liu, and Dawei Yin. 2019. Reinforcement learning to optimize long-term user engagement in recommender systems. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2810–2818.