Grant Agreement No.: 823783
Call: H2020-FETPROACT-2018-2020

Topic: H2020-FETPROACT-2018-01
Type of action: RIA

D4.2 Algorithms and Implementation of Diversity Aware Single user incentive design

Revision: v.1.0

Work package	WP 4
Task	Task Number T4.2
Due date	30/06/2021
Submission date	30/06/2021
Deliverable lead	Avi Segal
Version	1.0
Authors	Daniel Ben Zaken, Kobi Gal, Avi Segal (BGU)
Reviewers	Nardine Osman (CSIC)

Abstract	We describe our work of developing non-monetary incentives for increasing motivation and improving productivity and engagement for single users. We elaborate on our research and implementations of harnessing a recommendation-based method focusing on increasing engagement and satisfaction by matching tasks to users. Our recommendation methods are diversity aware by generating recommendations based on users' behavioural patterns as opposed to their fixed demographic characteristics. We extend this research by designing adaptive explanations based on user's characteristics and behaviour in the system. In our research we take a computational based approach, which uses participants' interaction histories in the system to predict their future responses to considered incentives. Our implementations and trials point to the potential of these approaches and to the promise of designing adaptive explanations for these environments. In this research we work with systems in a large-scale volunteer-based crowdsourcing domain which serves as a fertile ground for our experimentations in the wild.
Keywords	Incentive Design, Recommendation Systems, Adaptive Explanations

Document Revision History

Version	Date	Description of change	List of contributor(s)
V0.8	08/06/2021	Version sent to WP4 internal review	Daniel Ben Zaken, Avi Segal (BGU)
V0.9	14/06/2021	Version sent to WeNet internal review	Daniel Ben Zaken, Kobi Gal, Avi Segal (BGU)
V1.0	27/06/2021	Version after WeNet internal review	Daniel Ben Zaken, Kobi Gal, Avi Segal (BGU)

DISCLAIMER

The information, documentation and figures available in this deliverable are written by the "WeNet - The Internet of US" (WeNet) project's consortium under EC grant agreement 823783 and do not necessarily reflect the views of the European Commission.

The European Commission is not liable for any use that may be made of the information contained herein.

COPYRIGHT NOTICE

© 2019 - 2022 WeNet Consortium

Project co-funded by the European Commission in the H2020 Programme		
Nature of the deliverable:		**R**
Dissemination Level		
PU	Public, fully open, e.g. web	✔
CL	Classified, information as referred to in Commission Decision 2001/844/EC	
CO	Confidential to WeNet project and Commission Services	

** R: Document, report (excluding the periodic and final reports)*

DEM: Demonstrator, pilot, prototype, plan designs

DEC: Websites, patents filing, press & media actions, videos, etc.

OTHER: Software, technical diagram, etc.

EXECUTIVE SUMMARY

We report on our next step in developing non-monetary incentives for increasing motivation and improving productivity and engagement for single users in online systems. Specifically, we focus in this report on providing explanations for recommendations supplied to different user diversities based on their behaviour in an online system. We elaborate on our research and implementations of a recommendation-based method focusing on increasing engagement and satisfaction by matching tasks to users and on adapting explanations for each recommendation presented to users. We take a computational based approach, based on users' histories in the system.

Our recommendation-based work matches thousands of volunteers with Citizen Science projects in which people contribute online to scientific research. As a first step, we work with SciStarter, a leading citizen science portal on the web, which offers more than 3,000 projects and recruits volunteers through media and other organizations, bringing citizen science to people. We note that this work will be integrated into the WeNet system once enough usage data is available in WeNet. Given the sheer size of available projects in SciStarter, it faces two major challenges. First, how to find the right project, which best suits the user preferences and capabilities. This is deemed essential for keeping volunteers motivated and active. Second, once a project is recommended to a user, how to explain this recommendation to the user based on their characteristics and behaviour in the system. This user adapted explanation is considered necessary from both an ethical perspective as well as in order to increase adoption and contribution.

We develop a novel approach for supplying the right recommendation to volunteers based on their profile and logged activities. We use methods from the field of recommendation systems, including collaborative filtering methods and matrix factorization. We then develop a state-of-the-art approach for selecting the best explanation for each recommendation per user. This approach considers the ranked recommendation list of items created for each user, multiple available explanation types per item and the manner in which the recommendations and explanations will be presented to users. This research develops a new approach for personalized recommendations and explanations in the SciStarter ecosystem, in order to increase their engagement and motivation to contribute to citizen science platforms. We lay the foundation to enhance participants' motivation and learning in online systems by aligning interests with projects through AI recommendation and explanation tools.

We evaluated our explanation approach in the SciStarter system, based on data collected from thousands of users who previously participated in an online study involving AI based recommendation of new projects. In the online study, volunteers were randomly divided into different cohorts, which varied the recommendation algorithm that was used to generate suggested projects. We were able to show that: 1) our recommendation system was able to engage people in new projects that they had never tried before; 2) the recommendation system led to increased participation in SciStarter projects; and 3) the cohort of volunteers receiving recommendations created by the SVD algorithm (matrix

factorization) exhibited the highest levels of contributions to new projects, when compared to the other cohorts. This was the first study using AI based recommendation tools in large scale citizen science platforms.

We used the data from the online study to develop our explanation algorithm and test it in an offline setting. We show that the newly developed algorithm outperforms alternative approaches offline. We then worked with SciStarter to design and prepare an online study for testing the explanation algorithm in the wild. This study will launch in the coming month.

Transferability of the developed approaches to the WeNet system will be achieved by mapping items (which are projects in the case of SciStarter) to the activities offered in WeNet apps.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

Citizen science engages people in contributing to the science by collecting, categorizing, transcribing, or analyzing scientific data [4, 11, 5]. These platforms[1] offer thousands of different projects which advance scientific knowledge all around the world. With this wide variety of projects, finding the right project, which best suits the user preferences and capabilities, has become a major challenge [22, 6] and is essential for keeping the user motivated, active and satisfied. In this research we aimed to test how a personalized AI-based recommendation system can address this challenge and incentivise users, increase their motivation and satisfaction with the projects they choose to contribute to. We focus on two problems which complement each other: generating personalized recommendations and explaining these recommendations to an end user in an adaptive manner.

Recommendation systems have been used in other domains, such as e-commerce, news, social media, content aggregation and censorship systems [14, 8]. Personalized recommendations have been shown to constitute strong incentivising mechanisms in monetary based systems, where they are used and studied extensively [29]. Specifically, past work has shown that personalized recommendations are incentivising and leading to increased motivation, higher engagement levels, extended consumption level and increased purchasing behaviour [31,32]. Additionally, applying personalized recommendations in online educational settings has shown promising impact on the motivation and learning gains of target learners [30]. Thus, such mechanisms are deemed important in the context of a diversity focused social project such as WeNet, where non-monetary policies are to be used for incentivising users and achieving system goals.

In our past report we developed a new approach for personalized recommendations in citizen science platforms, in order to increase their engagement and motivation to contribute. Our recommendation system delivered personalized recommendations to signed-in users by recommending them with new projects based on their past history on the site and based on projects' characteristics. We applied this approach in the wild in the SciStarter platform (https://scistarter.org). SciStarter offers more than 3,000 projects and recruits volunteers through media and other organizations, bringing citizen science to people. Our results pointed to the potential of the proposed recommendation based approach and algorithms to incentivize users in voluntary non-monetary domains, such as the WeNet system.

In this report we extend our work by using data from our past study to develop explanation algorithms for recommendation systems. Specifically, we develop post hoc explanation approaches for black box recommendation algorithms which are hard to explain. This approach generates adaptive explanations while addressing different characteristics of the user's diverse behaviour in the system. We show that our proposed

[1] E.g zooniverse.org, scistarter.org

algorithm outperforms alternative approaches in offline settings. We now work with SciStarter to design and prepare an online study for testing our approaches in the wild.

2 PAST WORK

This work relates to past work in citizen science research as well as recommendation and explanation systems. We first discuss work from the field of citizen science, and the use of AI in this domain in order to increase users' engagement. Then we describe in detail the field of recommendation systems and their applications. Finally we present relevant past work on explanations for recommendation systems.

2.1 CITIZEN SCIENCE – MOTIVATION AND LEVEL OF ENGAGEMENT

Online participation in citizen science projects has become very common [18]. Yet, most of the contributions rely on a very small proportion of participants. This group of participants is highly engaged with the projects and makes thousands of tasks. However, in most citizen science projects, the majority of participants carry out only a few tasks. Many researches have explored the incentives and motivations of participants in order to increase participants' engagement. Kragh et al. [15] claimed that participants in citizen science projects are motivated by personal interest and desire to learn something new, as well as by the desire to volunteer and wish to contribute to science. Raddic et al. [20] support the above claim and discuss that participants' engagement is mostly originated in pure interest in the project topic, such as astronomy. Arazy et al. [18] have explored this by separating the question to the quantity of contribution and quality of contributions. They showed that quantity of contribution is mostly determined by the user interest in the project and by social norms. In contrast, the quality of contribution is determined by understanding the importance of the task and by the user's reputation. In our work, we aim to increase both the quantity and quality of contributions, based on the described factors.

Significant prior work was done in order to increase participants engagement, which is influenced by the discussed motives. Segal et al. [27] have developed an intelligent approach which combines model-based reinforcement learning with off-line policy evaluation in order to generate intervention policies which significantly increase users' contributions. Laut et al. [17] have demonstrated how participants are affected by virtual peers and showed that participants' contribution could be enhanced through the presence of virtual peers.

Ponciano et al. [19] characterized volunteers' task execution patterns across projects and showed that volunteers tend to explore multiple projects in citizen science platforms, but they perform tasks regularly in just a few of them. They have also shown that volunteers recruited from other projects on the platform tend to get more engaged than those recruited outside the platform. This finding is a great incentive to increase user engagement through platform instead of targeted projects like we do in our research.

In this research, we attempt to enhance participant engagement to citizen science projects by recommending the user projects which best suit the user preferences and capabilities.

2.2 RECOMMENDATION SYSTEMS

Recommendation systems aim to provide the most relevant items to users by learning their choices and producing the results that correlates to their interests and needs [21]. Generally, recommendation systems are based on two main approaches: a content-based approach and collaborative filtering approach.

The content-based strategy [3] is based on shared items and users' characteristics. For example, item's characteristics - a project in the SciStarter case - can include the project task, location and targeted age group. The user is recommended with items according to these characteristics. Alternatively, collaborative filtering (CF) strategy [25] is based only on users' past activities, and recommendation is based on analyzed relations between users and items. Recommendations based on CF rely on different types of input. Preferably, explicit data, which is based on some sort of user-item rating. Generally, this type of input is not available in many domains. Therefore, implicit data is taken into consideration, where user-item interactions are inferred from user behaviour, such as participation, clicks, search patterns etc. A good discussion on collaborative filtering based implicit feedback is given by Hu. et al. [13]. Hu. et al. discuss the unique characteristics of implicit data, which justify the replacement of explicit based algorithms. In the educational domains, other works also tried to increase users' engagement and participation by personalized recommendations. Labarthe et al. [7] built a recommender system that recommends relevant and rich-potential contacts based on user profile and activities. They have shown that by recommending this list of contacts, students were much more likely to persist and engage in MOOCs. A subsequent work of Dwivedi et al. [16] has developed a recommender system that recommends online courses to students based on their grades in other subjects. This recommender was based on collaborative filtering techniques and particularly item based recommendations.

Some other works that concern user engagement with recommendation systems have shown how early intervention significantly increases user engagement [9]. Wu et al. [28], have shown how tracking user's clicks and return behaviour succeeds to increase user engagement with the recommendation system. They have formulated the optimization of long-term user engagement as a sequential decision making problem, where a recommendation is based on both the estimated immediate user click and the expected clicks results from the users' future return.

All the systems described above are designed for different goals. In this work, we introduce a recommendation system with the goal of increasing user engagement and satisfaction with a new focus on citizen science projects. This system will take into consideration the special factors relevant to a volunteer based domain (such as SciStarter, WeNet and the likes), and by this, we develop a new strategy for keeping the user engaged with citizen science projects.

2.3 EXPLANATIONS FOR RECOMMENDATION SYSTEMS

Explanations introduced into recommendation systems can improve their prediction accuracy as well as increase recommendations acceptance by users. McAuley et al. [43] extracted topics in users' reviewed text which were used to explain the users' variation in rating. This improved the product rating prediction accuracy as well as produced a valid explanation. Bilgic et al. has shown that an effective explanation can help the user evaluate the quality of suggested items according to their own preferences. This increases the likelihood that users discard irrelevant options while helping them to recognize good ones [44]. Explanations can also improve the transparency, persuasiveness, effectiveness, trustworthiness, and user satisfaction from the recommender system [45].

Various forms of explanations have been explored in prior work, including sentences, word clouds, as well as different kinds of visualizations [46].

One way to categorize explainable recommendation research is by the model used for generating explanations. Abdollahi et al. [47] proposed explainability constrained Matrix Factorization technique that computes the top-n recommendation list from items that are explainable. The model produces neighbor-style explanations (user-based and item-based collaborative filtering) for items and users that meet threshold conditions. The threshold constraints lead to a lack of explanations for certain items and can harm the model predictions.

Another approach to provide explanations for complex models as Matrix Factorisation and other non explainable recommendation systems is a post hoc method. This approach seeks to interpret a black-box model after it has been trained, by extracting explanations from the output of the model while no changes are made to the black-box model itself [48]. Musto et al. [49] used aspect extraction of users' product reviews with natural language processing and sentiment analysis techniques. They generated automatic natural language explanations using aspects extracted from the reviews text. Note that this work generates explanations with the help of reviews written for items, which may not always be available.

The most relevant paper to our study is [37] which used post hoc explanations for Matrix Factorization based recommendations. Their method used common explanations such as neighbor style explanations and content-based explanations. Given the output of the recommender, the authors run a set of explainable recommendation algorithms which provide a score for the items recommended by the black box Matrix Factorization recommender. When this score is sufficiently high, it means that the explainable recommender agrees with the black box recommender and its explanation can be used. As stated in [37], their approach depends on the manually tuned thresholds for the acceptability of the explanations for a given Item. These thresholds are application dependent and hard to finetune. We are avoiding this constraint altogether by normalizing the scores of all models and by choosing explanations based on a ranking criteria. We are also extending [37] by developing a wider range of explanations and by targeting a large scale experiment in the wild.

3 RECOMMENDATION EXPERIMENT DESIGN

We focus on whether intelligent recommendation of citizen science projects will result in increased engagement and improved contributions as compared to existing SciStarter tools. We then move to consider different explanation methods for the presented recommendations and their impact on contributions and engagement. To address recommendations and explanations, we use and compare several recommendation algorithms that are based on existing state-of-the-art tools, adapted to the citizen science domain.

3.1 INPUT DATA

During the generation of personalized project recommendations, we consider a number of data sources. (1) Interactions with affiliate projects: data generated as a result of users' activities with affiliate projects, e.g., joining a project, making a contribution to a project or participating in a project. An affiliate project is one that uses a specific API to report back to SciStarter each time a logged in SciStarter user has contributed data or analyzed data on that project's website or app. (2) Direct interactions with projects on Scistarter's website; any type of click a user performs in Scistarter is recorded. The clicks that are considered as an interaction with a project are any clicks related to the project - could be searching the project, filling a form about the project, etc. The data is represented as a binary matrix R[U * I], where U is the number of users, and I is the number of projects. R[i, j] = 1 indicates that user i had some interaction with project j, and R[i, j] = 0 otherwise. (Interaction is either an interaction with affiliate project or interaction with projects on SciStarter). Consequently, our system is a top-N recommender system rather than a predictor.

3.2 RECOMMENDATION ALGORITHMS

We design four different algorithms: CF user based, CF Item based, Matrix Factorization - SVD, Popularity based.

3.2.1 User Based Collaborative Filtering

In this algorithm, the recommendation is based on user similarities [25]. The ranking of a project for a target user is computed by comparing users who have done similar projects. For example, the algorithm recommends the iNaturalist project (an online biodiversity project offered in SciStarter) to user A, because similar users to the target user have contributed to this project.

We use KNN algorithm to find similar users. We choose K = 100, which we found to be the optimal k for this problem. At first, we encounter a problem by choosing this K, since in our data, we have users who act exactly the same. Therefore, when we try to find the most similar projects of k similar users to user i, we get exactly the projects user i has already participated in. For this reason, we increase the value of K to K + 100 in each iteration, until we have a sufficient variety of projects from similar users (where similar is now a more broader definition).

3.2.2 Item Based Collaborative Filtering

In this algorithm, the recommendation is based on projects similarities [25]. The algorithm ranks a project for the target user by comparing its similarity to other projects. For example, many users engaging in CoCoRaHS (an outdoor precipitation monitoring project) also sustained engagement in StallCatchers (an online project designed to accelerate Alzheimer's research).

Therefore, Stall Catchers will be recommended to a user who engages in CoCoRaHS if they have not yet already engaged in Stall Catchers. The similarity is calculated by cosine similarity.

3.2.3 Matrix Factorization - SVD

This algorithm predicts the user's rating for each project and projects with the highest predicted ratings are recommended to the user [24].

SVD states that any matrix R can be factorized as: $R = USV^{T}$.
This algorithm is used in recommendation system in order to find the multiplication of the three matrices U, S, V^{T}, to estimate the original matrix R and hence, to predict the missing values in the matrix. (The matrix R includes missing values since users did not participate in all projects, and therefore we would like to estimate how much will a user like an unseen project). In the settings of recommendation system, the matrix U is a left singular matrix, representing the relationship between users and latent factors. S is a diagonal matrix describing the strength of each latent factor, while V^{T} is a right singular matrix, indicating the similarity between items and latent factors. Latent factors describe a property or concept that a user or an item have. For example, a latent factor can refer to where the project takes place (indoors or outdoors, online or offline, etc.). SVD decreases the dimension of the utility matrix R by extracting its latent factors. It maps each user and item into a latent space with r dimensions and with this, we can better understand the relationship between users and projects, as they become directly comparable.

For example, the project Asteroid Mappers (an online project designed to identify craters on the asteroid Vesta) is recommended to user A, since this project has the highest predicted rating than all other projects, for this target user. It has the highest predicted rating mainly because the user has participated in the past in projects Moon Mappers (an online project designed to perform science tasks on Mars) and StallCatchers (described above), which are similar to project Asteroid Mappers and also because project Asteroid Mappers best describes the correlation of the target user with other users.

3.2.4 Popularity Based

In this algorithm, all projects are sorted by popularity level [1]. Popularity is measured by the number of users who contribute to the project. For example, the project Globe at night (a project designed to measure the night sky brightness) is one of the most popular projects in SciStarter, and many users participate in it. Therefore, this project is recommended for users assigned to this algorithm.

3.3 OFFLINE EVALUATION

In this section, we evaluate the different recommendation algorithms on historical data from SciStarter. We used data that included logged-in SciStarter-users, who interacted with at least two projects. (Interaction is defined above). The experiment included 6353 users, where each user was assigned to each one of the algorithms.

Evaluation Methodology

We chronologically split the data into train, and test sets such that 10% of the latest interactions from each user are selected for the test set and the remaining 90% of the interactions are used for the train set.

We evaluate the top-n recommendation result. We use two of the most common top-n metrics: Precision and Recall. Precision: the proportion of recommended projects that are relevant (Relevant projects are the projects that the user eventually interacted with, either by clicking and/or participating), Recall: proportions of relevant projects that were recommended. In addition, we also examined a new metric named Refined Precision, which was suggested in [10]: out of the rejected items (recommended but not interacted), Refined Precision describes the sum of max similarities to interacted projects. Precision, Recall and Refined Precision were measured at 3, where 3 is the number of recommended projects. The number 3 was chosen since SciStarter presents three recommended projects by default.

3.4 ONLINE SETTING

After evaluating the algorithms in an offline setting, we also wanted to test their performance online. We performed three online experiments. The first experiment took place between 16.9.2019 and 11.10.2019. The second experiment took place between 11.10.2019 - 2.12.2019 and included recommendations based on user's physical location. In the first experiment, users were provided with recommendations which were not necessarily close to the user's physical location. Therefore, it was not possible for users to participate in such projects. For example, the project "Vermont Atlas of Life on iNaturalist" takes place in Vermont, U.S, and users from Europe could not participate in this project. Hence, we restricted our algorithms to recommend only projects that are physically close to the user's physical location. By taking into account the IP address with which the user enters SciStarter, we could infer her landmark, and use it as a restriction for the recommendations. For each such user, we filtered the recommended projects by this restriction and recommended on the top n projects which are in the same region as the user or projects which are taken online. Lastly, the third experiment took place between 2.12.2019 - 17.2.2020. In this experiment, we initialized users' allocation to cohorts from fresh, since the clicks stream data which our algorithms were based on, was unreliable before. In this paper, we will describe only the last experiment.

3.4.1 Participants

A total of 1331 logged-in users have participated in the online experiment, which was conducted between December 2019 and February 2020. Out of the 1331 users, only 131 users interacted with at least two projects. These users were assigned equally into five cohorts, while the rest of the users got the default recommendation of the Popularity algorithm. The users allocation to cohorts was round robin allocation since we wanted the cohorts to be of the same size. Each cohort was provided with recommendations based on the different algorithm: (1) Item-based Collaborative filtering, (2) User- based Collaborative filtering, (3) Matrix Factorization - SVD,(4) Popularity based, (5) Baseline, where 'Baseline' is a cohort which did not get any recommendation from the system and instead, provides the user with Scistarter's promoted projects as an alternative to the recommendations.

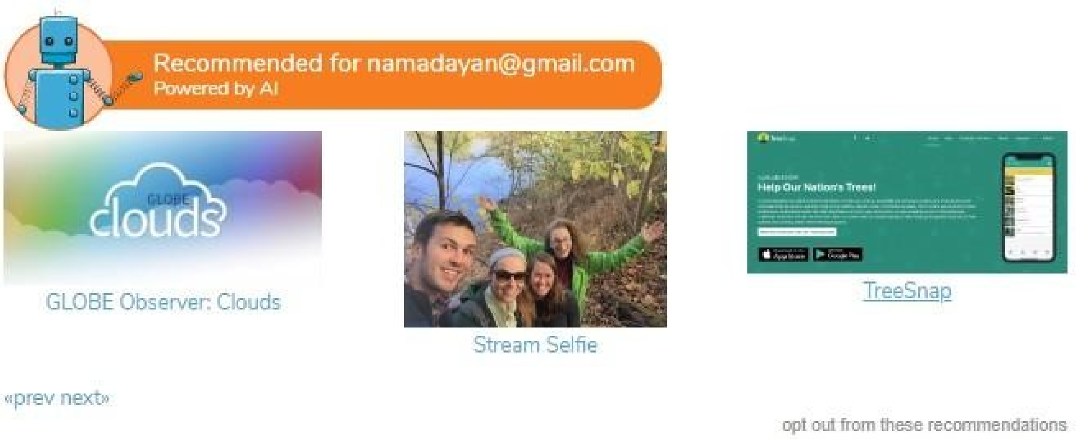

FIGURE 1: SCREENSHOT OF RECOMMENDATION TOOL FROM ONLINE EXPERIMENT

3.4.2 Evaluation Methodology

In our system, we use live evaluation via A-B testing to evaluate the recommendation system performance. Users are assigned to one of the five cohorts described above, where the baseline cohort is the control group. We measure precision, recall and refined precision metrics in the online experiment as well. However, in our domain, the number of recommended items presented to the user changes. Users are presented with three projects and can choose to get more recommendations. Thus, instead of calculating precision, recall and refined precision at 3, where 3 is the number of recommended projects, we calculate these metrics at the actual number of recommendation each user is exposed to.

In order to evaluate recommendation quality in the online experiment, we use a number of different metrics:

- RecE - number of users' actions of clicking on a recommended project and/or interacting with it later.
- NoRecE - number of users' actions where the user did not click on a recommended project but did interacted with the project later.
- NoRecNoE - number of users' interactions with projects that were not recommended to her, and

- ToolE - number of any user’s engagement with the recommendation tool (e.g. clicking the project, clicking on the project’s image or ’next’, ’prev’ buttons).

With these metrics, we evaluate system changes on live traffic and also track the system performance at an ongoing basis.

4 RECOMMENDATION RESULTS

4.1 OFFLINE RESULTS

Table 1 contains results of off-line evaluation for the 4 examined algorithms, where the described metrics are precision, recall and refined precision at 3. User-based collaborative filtering and SVD outperform the other algorithms. Popularity algorithm generates the lowest performance on precision and recall.

Algorithm	Precision@3	Recall @3	Refined Precision@3
Item Based Collaborative filtering	4.6%	43.2%	41.7%
User Based Collaborative filtering	22.7%	68.6%	53.7%
Popularity	11.1%	33.3%	44.5%
SVD	18.6%	54.9%	45.0%

TABLE 1: OFF-LINE RESULTS FOR RECOMMENDATIONS ALGORITHM TYPES.

4.2 ONLINE RESULTS

The recommendation tool was active on SciStarter for over 41 days and gave the following results:
26-27 users were allocated to each cohort, and no user has chosen to opt out from the experiment. A total of 624 clicks on recommendations were recorded, which is about 5.6% Click Through Rate (CTR) (including users with less than 2 interactions). Table 2 presents the described recommendation quality metrics for each examined algorithm. We can see that SVD algorithm has the best performance in all metrics.

Algorithm	#Users	RecE	NoRecE	NoRecNoE	ToolE	Precision	Recall	Refined Precision
Item Based Collaborative filtering	25	13	3	56	25	2.9%	8.8%	25.3%
User Based Collaborative filtering	25	36	6	37	72	11.3%	28.2%	35.4%
Popularity	25	38	1	29	99	9.1%	22.5%	36.7%
SVD	25	54	8	64	124	11.4%	24.4%	37.8%
Baseline	25	15	1	29	29	9.3%	24.2%	37.0%

TABLE 2: ON-LINE RESULTS FOR RECOMMENDATION ALGORITHM TYPES

Some correlation exists between the offline and the online experiments. In both experiments SVD and User based collaborative filtering were the best algorithms. However, the performance of Item based collaborative filtering algorithm was significantly lower in the online experiment as compared to the offline experiment, while popularity algorithm proves itself better in the online experiment. We can see that 3 out of 4 algorithms (User based collaborative filtering, SVD and popularity) are better than the baseline algorithms in all tested metrics. Users are much more engaged with SciStarter and the projects as compared to the baseline approach.

We can reason SVD's good performance according to the literature, where SVD is considered as a leading algorithm in the domain of recommendation systems [12, 23]. In addition, in this particular work, SVD generates recommendations that are more rare (less familiar to users) than other algorithms do. A user study we have performed has shown that users are more interested in projects they have never heard about: "I did not click on either project because I have looked at both projects (several times) previously", "I am more interested in projects I didn't know exists before". This finding is another reason that explains why SVD performs better than the other algorithms. We can demonstrate the difference in the recommendations of SVD, which recommends more rare projects than other algorithms.

- User A, who is assigned to User based collaborative filtering algorithm receives the recommended projects: [167,734,2992], which their popularity scores are [11,16,19] respectively.
- User B, who is assigned to Item based collaborative filtering algorithm receives the recommended projects: [2992,83,114], which their popularity scores are [19,12,25] respectively.
- User C, who is assigned to SVD algorithm receives the recommended projects: [890,671,87], which their popularity scores are [3,1,6] respectively.

We can see that SVD recommends less popular projects than other algorithms.

5 RECOMMENDATIONS USER STUDY

In order to learn what are the users' opinions about the recommendations, and their level of satisfaction, we conducted a survey with SciStarter's users. Our survey was sent to all SciStarter community users.

One hundred and thirty eight users have filled the survey, where each user was asked about the recommendations presented to him by the algorithm he was assigned to. The survey included questions about users' overall satisfaction with the recommendation tool as well as questions about their pattern of behavior before and after the recommendations. See Appendix A for the survey details. The majority of users (75%) were very satisfied with the recommendation tool and claimed that the recommendations were suitable for their personal interests and goals. The majority of users (54%) reported they have clicked on the recommendations and visited the project's site (Figure 2), while only 8.8% of users did not click the recommendation or visited the project site, users who were not familiar with the recommended projects before, clicked more on the recommendations, as well as users who previously performed a contribution to a project.

Users who did not click on the recommendations can be divided into 3 main themes: (1) Users who don't have the time right now or will click the project in the future. (2) Users who feel that the recommendations are not suitable for their skills and materials: "Seemed out of my league", "I didn't have the materials to participate". This behaviour was also discussed in [26], and was named "classification anxiety". (3) Users who feel that the recommendations are not suitable for their interests: "No interest in stall catchers", "The photos and title didn't perfectly match what I am looking for". Not many users have also performed a contribution to the recommended projects (only 15%), but they reported their desire to do so in the future. About 93.4% of users reported they would use the AI recommendations tool in the future, while 92% of them are familiar with the search engine.

The survey demonstrated how powerful the impact of AI is. We got some very enthusiastic feedback on the recommendations: "I am very impressed by the new Artificial Intelligence feature from SciStarter! Your AI feature shows me example projects that I didn't know before exist", "I like how personalized recommendations are made for citizen science users", etc.

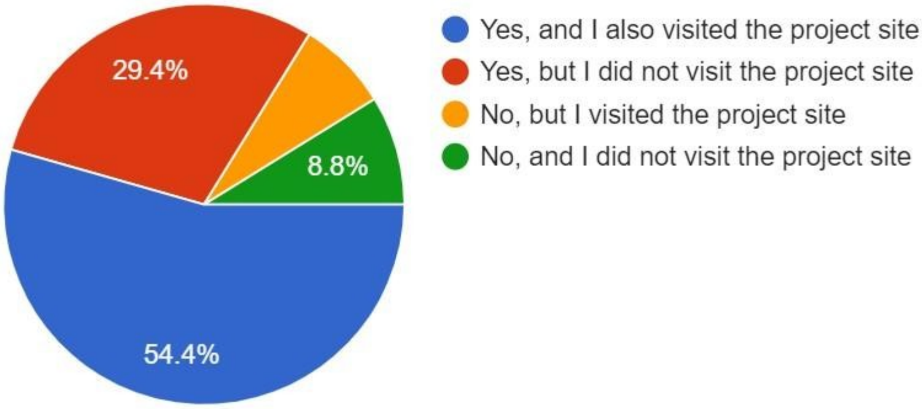

FIGURE 2: DID YOU CLICK ON ONE (OR MORE) OF THE RECOMMENDED PROJECTS?

6 USER AWARE EXPLANATIONS

We now move to describe our explanation approach. We strive to develop an explanation system that will use the recommendations generated for each user, provide different types of explanations while also considering the presentation constraints of the system. Specifically, the algorithm should address the following two challenges: (1) generate an explanation for every <user, project> recommendation tuple based on all historical information available on the user in the system, i.e. while being aware of the user historical data and behaviour in the system, and (2) organize recommendations and explanations in groups prior to presentation to users. This second challenge is a requirement presented by our SciStarter partner and it follows similar approaches for explanation presentation in recommendations engines. E.g the standard recommendations in Netflix are group based, aggregated per different explanation types. E.g, some series are recommended to you "because you liked Bridgerton" and these are grouped and presented together, in separation from series recommended "because you may like movies made in Italy". Any explanation algorithm developed will need to take such grouping into consideration.

6.1 IMPROVED RECOMMENDATION SYSTEM

Our recommendation algorithms so far (sections 3 and 4) did not consider the content available for each project in the system, namely the project's title, description etc.
To address this gap, we extend a well known Matrix factorization(MF) BPR algorithm [29, 34] with a content-based component to create a hybrid BPR-Content algorithm. This extended algorithm works as follows:

A content-based recommendation is generated based on several features extracted for each project. Specifically, these include the duration of each contribution activity (seconds), the project's platform (indoor/outdoor), the project's topics (textual values), and the project's tags (textual values). These features are defined by the projects' owner and pre-processed by us. The features are then encoded by one-hot encoding and cosine similarity [35] between feature pairs is calculated by measuring the angle between vectors. For binary settings, the cosine similarity may be computed using:

$$sim(p_1, p_2) = \frac{|F_{p_1} \cap F_{p_2}|}{|F_{p_1}| \cdot |F_{p_2}|}$$

where F_p is the set of features that projects p has.
We then compute a score for item i that user u has not yet interacted with:

Co-funded by the Horizon 2020 Framework Programme of the European Union

$$\hat{r}_{u,i} = \sum_{p' \in P_u} sim(p, p')$$

The final Hybrid recommendation model combines the content recommendation score and the BPR recommendation score for each <user, item> tuple. Specifically, it computes a weighted average of both scores with weights of 0.8 and 0.2 for BPR and content, respectively. These weights were optimized to achieve a maximal precision@3 metric on the research dataset, which is the number of recommendations that the users receive on the SciStarter home page.

6.2 OFFLINE ANALYSIS OF HYBRID MODEL

We compared our content based algorithm with the algorithms presented previously. Specifically, we measured hit rate and precision as defined earlier. Figure 3 presents both metrics as computed on the latest dataset provided to us by SciStarted.

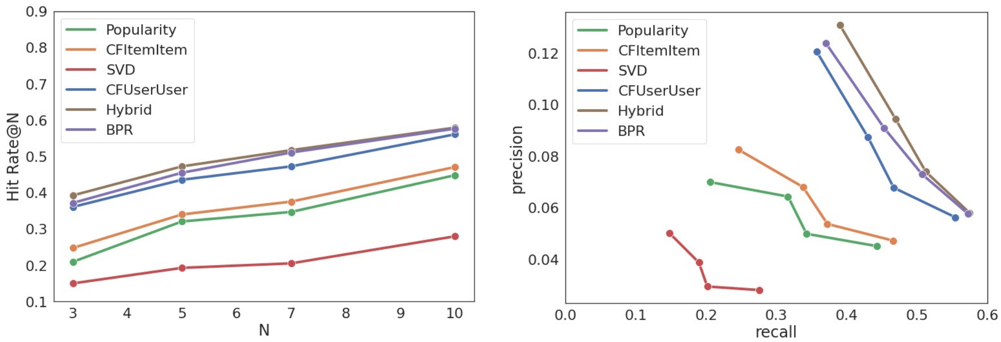

FIGURE 3: HIT RATE @N RECOMMENDATIONS ON TEST DATA (LEFT); PRECISION ON TEST DATA (RIGHT)

As can be seen from the Figure, the hybrid model outperforms all alternative algorithms on both metrics.
We note that the hybrid algorithm was implemented using the LightFM library [36] with the BPR (Bayesian Personalised Ranking) loss function.

6.3 DESIGNING EXPLANATIONS

The BPR matrix factorization approach uses latent features to generate project recommendations for each user. As such, there is no straightforward way to create meaningful explanations for the projects recommended by our hybrid recommendation model. To address this challenge, we instead take a post-hoc approach [37]. Our Post-hoc approach takes the following high level steps: (1) first, a target recommendation list

is generated using a (hard to explain) recommendation algorithm (2) then, alternative recommendation lists are generated using easy to explain alternative algorithms (e.g. item based collaborative filtering, feature based recommendation) (3) Finally, items in the target recommendation list are explained by using explanations from one of the alternative algorithms, based on some preference criteria. Such a criteria may be (e.g.): “explain this item based on the alternative algorithm which gave the highest score to the item among all alternative algorithms”.
Thus, we use several explainable recommendation models to explain each item recommended by our hybrid model. We now move to describe the different explainable models we use.

6.3.1 Topics Associations Model

Association Rule Mining is a data mining technique used to identify relationships between categorical items in a collection. An example of an association rule is the implication {Health & Medicine ⇒ Cell & Molecular} where Health & Medicine are the antecedent of the rule and Cell & Molecular is the consequent of the rule.

In this explanation type we use associations to explain recommendations generated for users. Specifically, newly recommended projects are associated with projects the users already contributed to. For this type we use the Apriori algorithm [38] to mine association rules on projects' topics from the users’ interactions data. We consider association with support higher than 0.01. This value was set empirically to ensure interesting rules can be generated for topics in the long-tail with low support (as suggested by [39]). The output of this technique is a set of antecedent and consequent topics and the explanations are generated according to antecedent topics of the user’s past interactions.

6.3.2 Feature Based Model

For this explanation type we built recommendations based on textual features available for each project (e.g. the description of the project, its topic etc.). Features were encoded by one-hot encoding and the cosine similarity between projects is calculated for vector pairs. Explanations are then generated based on feature similarity between projects the user has already visited and new projects for this user.

6.3.3 CF-ItemItem Model

In item-based K-Nearest-Neighbor, we compute for each project a neighbourhood of other projects that similar users have interacted with [25]. For example, our data shows that 83% of users who have contributed to the project Never-Home-Alone (a project surveying wildlife in the home) have also contributed to project iNaturalist. Therefore we

Co-funded by the Horizon 2020 Framework Programme of the European Union

may recommend iNaturalist for a user who has already contributed to Never-Home-Alone.
Again, we require a similarity metric between items. The cosine similarity for items can be computed using:

$$sim(i_1, i_2) = \frac{|U_{i_1} \cap U_{i_2}|}{|U_{i_1}| \cdot |U_{i_2}|}$$

where U_i is the set of users who have contributed to project i.
We then compute a score for item i that user u has not yet interacted with:

$$\hat{r}_{u,i} = \sum_{i' \in I_u} sim(i, i')$$

Explanations based on this model will then explain a new project using its consumption by users which also consumed another project which was also consumed by the target user.

6.3.4 CF-UserUser Model

Collaborative filtering assumes that users with a history of contributing to the same projects would prefer to contribute to the same projects in the future. A user-based KNN collaborative filtering algorithm [40], identifies for each user u a set of k “neighbors" (users with similar histories as u, in that they interacted with the same projects). Then, we can recommend to u new projects that other users in her “neighborhood" have interacted with. For example, if both users u_1 and u_2 have contributed in the past to projects CoCoRaHS and Globe at Night, and u_2 has also contributed to project Stall Catchers, then Stall Catchers may be a suitable recommendation for u_1. To determine whether users belong to the same neighborhood, we need to measure how similar they are in their past interactions with projects.

6.3.5 Popularity Model

Popularity explanations will state that a specific project is recommended due to its popularity. Thus, the popularity list can be generated by ranking all projects by the number of users who contribute to them.

6.4 EXPLANATION GENERATION ALGORITHM

Algorithm 1 presents the explanation algorithm developed as part of this project. The post hoc explanation approach received k recommended projects (generated by the hybrid BPR-Content algorithm) and returns the most suitable explanation for each recommended project for each user.

The algorithm further prioritise certain explanation types over others: specifically, Previous studies have shown that most users prefer content-based related explanations [37] compared to “neighborhood-style” explanations for collaborative filtering [41]. Moreover, rule mining approaches (such as association rules), have advantages for explainable recommendations, since in many cases they can generate very straightforward explanations for users [42]. Our hypothesis is then that the users will be more satisfied with topics based associations explanations. Therefore, our algorithm seeks to explain recommended projects in the following order: (1) first by the topics-associations rules model (2) then by either feature based or neighborhood based algorithms, according to higher recommendation score (3) then by popularity (4) and lastly, if any other explanation is not available, by a general generic explanation message.

Algorithm 1: Assign explanations types to Hybrid recommendations

```
Input: Ordered dictionary of K recommended projects ordered by rank and their empty
       explanations , Set of K projects from each explainable recommendations models.
Result: Dictionary of recommendations and their explanation
R = K recommendations from non explainable Hybrid model
ExplainableProjectsScores = K recommendations and their normalize scores from
  each model ∈ [ContentFeatureBase, CFItemItem, CFUserUser]
TopicsAssociationsRoles = K recommendations from Topics associations rules model
Popularity = K recommendations from Popularity model
for project ∈ R do
    if project ∈ TopicsAssociationsRoles then
        project.Explanation = TopicsAssociationsRoles.Explanation
    else
        if project ∈ ExplainableProjectsScores then
            project.Explanation =
              max_{Explanation} Score(project) ∈ ExplainableProjectsScores
        else
            if project ∈ Popularity then
                project.Explanation = Popularity.Explanation;
            else
                project.Explanation = General.Explanation
            end
        end
    end
```

© 2019-2022 WENET

Co-funded by the Horizon 2020
Framework Programme of the European Union

6.5 GROUPING AND REORDERING OF RECOMMENDATIONS AND EXPLANATIONS

As described above, our method assigns an explanation type for each user's recommendation. Thus, the overall list of projects recommended to a specific user will typically include several explanation types. For example, a user may be recommended with three projects: (1) iNaturalist, (2) never-Home-Alone, and (3) Globe at Night in this order. These best explanations for these projects for this user may be by Content-Features, Association rules, and again Content-Features, respectively. As explained earlier, projects need to be organized into groups of explanation types. This is due to a requirement from our SciStarter partner and in accordance with explanation grouping practices in other recommendation services (e.g. Netflix). Such grouping may also change the ranking of the recommended projects presented to users. E.g, in the previous example, if the system prefers to have the largest group first, it will first present the two projects explained by the Content-Features explanation type, thus now ranking "Globe at Night" higher than "iNaturalist". Alternatively, if the system first presents the Association rules explanation type, based on our hypothesis about users' explanations preferences, then the "Globe at Night" project will be presented first followed by the two other projects. While improving the presentation to users, this reordering of project recommendations might harm the quality of recommendations for users since the ordering is now different from the optimization performed by the ranking algorithm. To analyze this impact and decide on the grouping policy for our domain, we tested several grouping methods and their effect on the recommendation ranks as described in the following section.

6.5.1 Re-order by Leader

In this grouping approach the order of the explanation groups is kept as close as possible to the original recommendation order. For each K-ranked recommendations and their explanation types, we are grouping the explanation types and ordering them according to the rank of the first project of each group.

6.5.2 Re-order by max size

In this grouping approach, the larger groups of explanations are present first. For each K-ranked recommendations and their explanation types we are grouping the explanation types and ordering them according to the size of the groups.

6.5.3 Re-order by explanation priority

As described earlier, according to existing literature, collaborative filtering approaches are less intuitive to explain compared with content-based algorithms.
Previous studies show [37] that most users prefer content-based explanations over collaborative filtering explanations or popularity-based explanations.
Thus, in this grouping method for each K-ranked recommendations and their explanations types we are grouping the explanation types and ordering the recommendations according to the following explanation priority:

1. Topics Association rules
2. Content Feature-based
3. Collaborative Filtering ItemItem
4. Collaborative Filtering UserUser
5. Non-personalized popularity based
6. Non-personalized general message

We evaluate the impact of the various grouping methods using the Precision@k, Recall@K, and HitRate@k for each method compared to the baseline - the BPR-Content Hybrid recommendation method. We aim to measure the negative impact on these metrics from every grouping (and reordering) method.

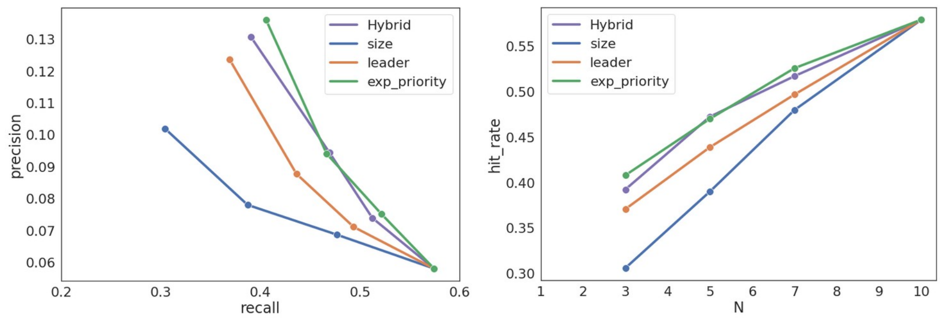

FIGURE 4: REORDER METHODS ON TEST DATA: PRECISION AND RECALL@N (LEFT); HITRATE @N (RIGHT)

Figure 4 presents the Precision and Recall metrics for K= 3, 5, 7, 10 recommendations (left) and the HitRate @N (right). Metrics were computed on a pre-split train/test of 90%/10% for every user's interactions. As seen in the figures, the grouping by the explanation priority methods outperforms all other grouping methods on all metrics. This is consistent with our hypothesis as to the preferred grouping method and provides an initial validation in an offline setting. Thus, this is indeed the grouping method which we will also use for our experimentation in the wild.

© 2019-2022 WENET

An interesting result is that for K=3, the number of projects recommended on the SciStarter homepage, the explanation priority grouping method outperforms the BPR-Content Hybrid baseline. This result needs further investigation, optimally with larger sample sizes. If persistent, one hypothesis here may be that users have latent preferences for items which are captured, at least partially, by the underlying explanation mechanism.

6.6 MAPPING EXPLANATION TYPES TO WENET'S DIVERSITY DIMENSIONS

The WeNet project addresses diversity through the lens of social practices. Social practices are routine behaviour like going to work, cooking and showering, which integrates different kinds of elements, such as bodily activities, material artefacts, skills, and associated meaning [50]. This view of diversity ventures away from a shallow notion of demographic differences between people (based on gender, race, geographic location, and the likes) to a deeper view, which considers multiple aspects of human existence and behaviour. Specifically, the literature defines the different aspects of social practices as follow:

- ***Material*** covers all physical aspects of the performance of practice encompassing objects, infrastructures, tools, hardware including the human body.
- ***Competence*** incorporates skills, know-how, (background) knowledge as well as social and relational skill which are required to perform the practice
- ***Meaning*** incorporates understanding, beliefs, values, norms, lifestyle, emotions, and social and symbolic significances

We now consider the different explanation types used in our algorithm and their relations to the Material, Competence and Meaning aspects. Table 3 presents the social practices aspects associated with each explanation type. As detailed in the table, all dynamic explanation types considered by our algorithm are covering social practices and are thus matching explanations to users based on their deep (and changing) behaviour in the system. Only if no explanation is obtained by any of these algorithms, our approach falls back to a popularity based or general based explanation. Indeed, in these two explanation types, the user's uniqueness and diversity is not taken into consideration, and a "one size fits all" solution is utilized as a last resort.

Explanation Type	Social Practice	Justification
Topics Association Rules	Material, Competence, Meaning	Projects associated together indicate **latent relations between topics**. These relations are the result of users' behaviour

		in the system which in turn are influenced by deep aspects of the users: the material they possess, their skills, and the meaning they attribute to different topics.
Content Feature Based	Competence, Meaning	Similar projects are identified using **textual relations between project metadata.** This metadata, available to the user, describes the expected effort of each project (competence) and its overall intentions (meaning).
Collaborative Filtering Item-Item	Material, Competence, Meaning	Similar projects are identified by **matching consumption patterns between users**. These patterns may be a result of any latent characteristic of users, not restricted to their demographic features.
Collaborative Filtering User-User	Material, Competence, Meaning	Similar projects are identified by **matching consumption patterns between users**. These patterns may be a result of any latent characteristic of users, not restricted to their demographic features.
Popularity Based	Shallow	Explanation is ignoring the user's unique behaviour in the system, using instead a fall back solution of "one size fits all" solution.
General Message	Shallow	Explanation is ignoring the user's unique behaviour in the system, using instead a general message.
Location Based	Shallow	Recommendation and explanations based on location are using a shallow demographic feature, failing to refer to the wider user's social practices.

Table 3: *Explanation Type Analysis - Diversity Dimensions*

© 2019-2022 WENET

6.7 TOWARDS DIVERSITY AWARE DXPLANATIONS IN THE WILD

We now report on the planned online experiment for our diversity aware explanation approach.

In this experiment, users who log on to SciStarter will be randomly assigned to one of two cohorts. One cohort will present project recommendations without explanations, and the second cohort will present project recommendations as well as explanations based on our described approach. The random allocation of users to cohorts will consider past covariates (e.g. past activity in the system) to decrease experiment bias to minimum. Each cohort will be provided with recommendations based on the BPR-Content hybrid model. The online experiment will address the following research questions:

1. What is the impact of explanations on users? Specifically:
    a. Are user who receive explanations contributing more to projects than users who do not receive explanations (but rather only recommendations)
    b. Do users who receive explanations press more on recommended projects than users who do not receive explanations (but rather only recommendations)
2. What is the opinion of users about explanations in a post-experiment study?

Both groups will be exposed to three recommended projects on the SciStarter home page and 20 projects on the recommendation page. Explanations will be presented by a header for each explanation group and by a tooltip for each recommended project - the question mark next to the project name. The tooltip will provide a more detailed explanation for the recommended project - see Figure 5.

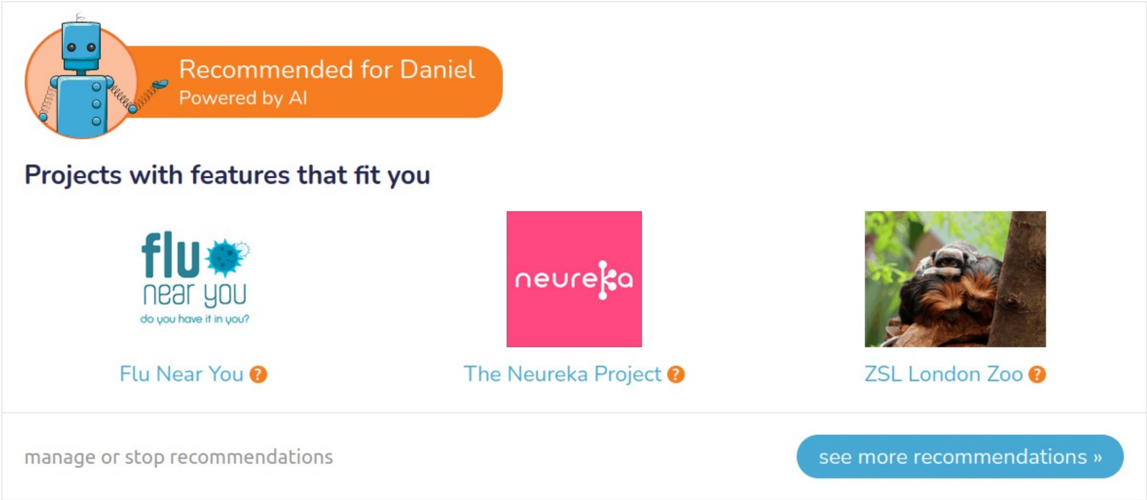

FIGURE 5: RECOMMENDED PROJECTS WITH HEADER ("PROJECTS WITH FEATURES THAT FIT YOU"

As seen in the figure, Daniel received three projects explained by the header “Projects with features that fit you”. These headers are generated based on the explanation type chosen for the presented projects. Table 4 presents the different explanation types and the headers and tooltip examples (detailed explanations) for each one of them.

Type	**Header**	**tooltip Example**
Topics Association Rules	Try projects with new topics	People who did Nature projects liked Forest projects.
Content Features	Projects with features that fit you	Nature Watch may fit your schedule. <or> You were interested in topic <topic> in the past
Items Similarity	Similar to projects that you like	People who liked Monarch Watch also liked Stall Catchers.
User Similarity	Others Like you like those projects	
Popularity	Hot on SciStarter	72% of SciStarter users love StallCatchers!
Location Base	Projects near you	
General	Try something new	

TABLE 4: AVALIBLE EXPLANATION TYPES

We are now working with SciStarter to finalize the details of the online experiment. A blog post updating the community of the upcoming experiment is due to be released shortly. Following this, we will move to introduce these algorithms into the WeNet solution based on data that will be generated by the WeNet community.

REFERENCES

[1] H. J. Ahn. Utilizing popularity characteristics for product recommendation. International Journal of Electronic Commerce, 11(2):59–80, 2006.

[2] M. Aristeidou, E. Scanlon, and M. Sharples. Profiles of engagement in online communities of citizen science participation. Computers in Human Behavior, 74:246–256, 2017.

[3] M. Balabanović and Y. Shoham. Fab: content-based, collaborative recommendation. Communications of the ACM, 40(3):66–72, 1997.

[4] R. Bonney, C. B. Cooper, J. Dickinson, S. Kelling, T. Phillips, K. V. Rosenberg, and J. Shirk. Citizen science: a developing tool for expanding science knowledge and scientific literacy. BioScience, 59(11):977–984, 2009.

[5] D. Brossard, B. Lewenstein, and R. Bonney. Scientific knowledge and attitude change: The impact of a citizen science project. International Journal of Science Education, 27(9):1099–1121,2005.

[6] H. K. Burgess, L. DeBey, H. Froehlich, N. Schmidt, E. J. Theobald, A. K. Ettinger, J. HilleRis- Lambers, J. Tewksbury, and J. K. Parrish. The science of citizen science: Exploring barriers to use as a primary research tool. Biological Conservation, 208:113–120, 2017.

[7] S. Dwivedi and V. K. Roshni. Recommender system for big data in education. In 2017 5th National Conference on E-Learning & E-Learning Technologies (ELELTECH), pages 1–4. IEEE, 2017.

[8] D. M. Fleder and K. Hosanagar. Recommender systems and their impact on sales diversity. In Proceedings of the 8th ACM conference on Electronic commerce, pages 192–199. ACM, 2007.

[9] J. Freyne, M. Jacovi, I. Guy, and W. Geyer. Increasing engagement through early recommender intervention. In Proceedings of the third ACM conference on Recommender systems, pages 85–92. ACM, 2009.

[10] S. Frumerman, G. Shani, B. Shapira, and O. Sar Shalom. Are all rejected recommendations equally bad?: Towards analysing rejected recommendations. In Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization, pages 157–165. ACM, 2019.

[11] C. Funk, J. Gottfried, and A. Mitchell. Science news and information today. Pew Research Center, 2017.

Co-funded by the Horizon 2020
Framework Programme of the European Union

[12] S. Gower. Netflix prize and svd, 2014.

[13] Y. Hu, Y. Koren, and C. Volinsky. Collaborative filtering for implicit feedback datasets. In 2008 Eighth IEEE International Conference on Data Mining, pages 263–272. Ieee, 2008.

[14] J. Itmazi and M. Gea. The recommendation systems: Types, domains and the ability us- age in learning management system. In Proceedings of the International Arab Conference on Information Technology (ACIT'2006), Yarmouk University, Jordan, 2006.

[15] G. Kragh. The motivations of volunteers in citizen science. environmental SCIENTIST, 25(2): 32–35, 2016.

[16] H. Labarthe, F. Bouchet, R. Bachelet, and K. Yacef. Does a peer recommender foster students' engagement in moocs?. International Educational Data Mining Society, 2016.

[17] J. Laut, F. Cappa, O. Nov, and M. Porfiri. Increasing citizen science contribution using a virtual peer. Journal of the Association for Information Science and Technology, 68(3):583–593, 2017.

[18] O. Nov, O. Arazy, and D. Anderson. Scientists@ home: what drives the quantity and quality of online citizen science participation? PloS one, 9(4):e90375, 2014.

[19] L. Ponciano and T. E. Pereira. Characterising volunteers' task execution patterns across projects on multi-project citizen science platforms. In Proceedings of the 18th Brazilian Symposium on Human Factors in Computing Systems, page 16. ACM, 2019.

[20] M. J. Raddick, G. Bracey, P. L. Gay, C. J. Lintott, P. Murray, K. Schawinski, A. S. Szalay, and J. Vandenberg. Galaxy zoo: Exploring the motivations of citizen science volunteers. arXiv preprint arXiv:0909.2925, 2009.

[21] F. Ricci, L. Rokach, and B. Shapira. Introduction to recommender systems handbook. In Recommender systems handbook, pages 1–35. Springer, 2011.

[22] F. Ricci, L. Rokach, and B. Shapira. Recommender systems: introduction and challenges. In Recommender systems handbook, pages 1–34. Springer, 2015.

[23] R. A. Sadek. Svd based image processing applications: state of the art, contributions and research challenges. arXiv preprint arXiv:1211.7102, 2012.

[24] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Incremental singular value decomposition algorithms for highly scalable recommender systems. In Fifth international conference on computer and information science, volume 1. Citeseer, 2002.

[25] J. B. Schafer, D. Frankowski, J. Herlocker, and S. Sen. Collaborative filtering recommender systems. In The adaptive web, pages 291–324. Springer, 2007.

[26] A. Segal, Y. Gal, R. J. Simpson, V. Victoria Homsy, M. Hartswood, K. R. Page, and M. Jirotka.

Improving productivity in citizen science through controlled intervention. In Proceedings of the 24th International Conference on World Wide Web, pages 331–337, 2015.

[27] A. Segal, K. Gal, E. Kamar, E. Horvitz, and G. Miller. Optimizing interventions via offline policy evaluation: Studies in citizen science. In Thirty-Second AAAI Conference on Artificial Intelligence, 2018.

[28] Q. Wu, H. Wang, L. Hong, and Y. Shi. Returning is believing: Optimizing long-term user engagement in recommender systems. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, pages 1927–1936. ACM, 2017.

[29] Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems. Computer. 2009 Aug 7;42(8):30-7.

[30] Drachsler H, Verbert K, Santos OC, Manouselis N. Panorama of recommender systems to support learning. InRecommender systems handbook 2015 (pp. 421-451). Springer, Boston, MA.

[31] Pathak B, Garfinkel R, Gopal RD, Venkatesan R, Yin F. Empirical analysis of the impact of recommender systems on sales. Journal of Management Information Systems. 2010 Oct 1;27(2):159-88.

[32] Gomez-Uribe CA, Hunt N. The netflix recommender system: Algorithms, business value, and innovation. ACM Transactions on Management Information Systems (TMIS). 2015 Dec 28;6(4):1-9.

[33] Schelenz L, Segal A, Gal K. Best Practices for Transparency in Machine Generated Personalization. arXiv preprint arXiv:2004.00935. 2020 Apr 2.

[34] Koren, Yehuda, and Robert Bell. "Advances in collaborative filtering." *Recommender systems handbook* (2015): 77-118.

[35] Breese, John S., David Heckerman, and Carl Kadie. "Empirical analysis of predictive algorithms for collaborative filtering." *arXiv preprint arXiv:1301.7363* (2013).

[36] Kula, Maciej. "Metadata embeddings for user and item cold-start recommendations." *arXiv preprint arXiv:1507.08439* (2015).

[37] Shmaryahu, Dorin, et al. "Post-hoc Explanations for Complex Model Recommendations using Simple Methods." (2020).

[38] Agarwal, Rakesh, and Ramakrishnan Srikant. "Fast algorithms for mining association rules." *Proc. of the 20th VLDB Conference*. Vol. 487. 1994

[39] Peake, Georgina, and Jun Wang. "Explanation mining: Post hoc interpretability of

© 2019-2022 WENET

Co-funded by the Horizon 2020
Framework Programme of the European Union

latent factor models for recommendation systems." *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2018.

[40] Ning, Xia, Christian Desrosiers, and George Karypis. "A comprehensive survey of neighborhood-based recommendation methods." Recommender systems handbook (2015): 37-76.

[41] Daher, Julie, Armelle Brun, and Anne Boyer. A review on explanations in recommender systems. Diss. LORIA-Université de Lorraine, 2017.

[42] Zhang, Yongfeng, and Xu Chen. "Explainable recommendation: A survey and new perspectives." *arXiv preprint arXiv:1804.11192* (2018).

[43] McAuley, Julian, and Jure Leskovec. "Hidden factors and hidden topics: understanding rating dimensions with review text." *Proceedings of the 7th ACM conference on Recommender systems*. 2013.

[44] Bilgic, Mustafa, and Raymond J. Mooney. "Explaining recommendations: Satisfaction vs. promotion." *Beyond Personalization Workshop, IUI*. Vol. 5. 2005.

[45] Ferwerda, B., Swelsen, K., & Yang, E. (2018). Explaining Content-Based Recommendations. New York, 1-24.

[46] Amal, Saeed, et al. "Enhancing explainability of social recommendation using 2D graphs and word cloud visualizations." *Proceedings of the 24th International Conference on Intelligent User Interfaces: Companion*. 2019.

[47] Abdollahi, Behnoush, and Olfa Nasraoui. "Using explainability for constrained matrix factorization." *Proceedings of the Eleventh ACM Conference on Recommender Systems*. 2017.

[48] Peake, Georgina, and Jun Wang. "Explanation mining: Post hoc interpretability of latent factor models for recommendation systems." *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2018.

[49] Musto, Cataldo, et al. "Generating post hoc review-based natural language justifications for recommender systems." *User Modeling and User-Adapted Interaction* (2020): 1-45.

[50] Reckwitz A. Toward a Theory of Social Practices: A Development in Culturalist Theorizing”. European Journal of Social Theory. 2002;5(2):243-263. doi:10.1177/13684310222225432

© 2019-2022 WENET

Co-funded by the Horizon 2020
Framework Programme of the European Union

APPENDIX A: QUESTIONNAIRE

Hello there, SciStarter community member!

We have recently released a tool providing you with personalized project recommendations. This tool uses AI technology to provide you with new project ideas that are best suitable for your enjoyment! You can see an example of the tool below.

You can help us out to evaluate the effectiveness of the recommendation tool. We will hand out $10 Amazon gift cards to the first 100 responders . You can read more about the recommendation tool at this link To be eligible for the gift card please do the following:
1. Log on to SciStarter with your username and take notice of your personalized recommendations.
2. Fill out the survey below when you are done with your SciStarter session.

----

1. Were you familiar with any of the projects that were recommended for you? *

○ Yes
○ No

2. Did you click to find out more information on one (or more) of the recommended projects? *

○ Yes, and I also visited the project site
○ Yes, but I decided not to visit the project site No, but I intend to visit the project site
○ Later
○ No, and I do not intend to visit the project site

3. If you did NOT want to click on any of the recommended projects, can you tell us why?

....

4. If you decided to visit any of the recommended projects, did you also make a contribution to that project?

○ Yes
○ No

5. How satisfied are you from the recommendation tool (1-Not satisfied at all; 5 – Very satisfied)? *

1 2 3 4 5

6. How suitable were the recommended projects for your personal interests and goals (1- not suitable at all; 5 – Very suitable)? *

1	2	3	4	5

7. Do you plan to use the AI recommendation tool in the future?

- ○ Yes
- ○ No

8. What are your main reasons for choosing to contribute to a project? *

- ○ The topic of the project interests me (topic could be physics, birds, insects,
- ○ education, etc.) The task of the project interests me (photography, classification, transcription, etc.)
- ○ The project is nearby
- ○ The project can be done online
- ○ The project is similar to projects I contributed to in the past
- ○ The project is well known and popular
- ○ Other

9. How many total projects have you contributed to in SciStarter? *

- ○ I didn't contribute to any project
- ○ I contributed to 1-2 projects
- ○ I contributed to more than 2 projects

10. Have you ever used the SciStarter project search tool? *

- ○ Yes
- ○ No

11. Please let us know about any other comments and ideas you have regarding project recommendations in SciStarter. Your ideas are appreciated!

....

Co-funded by the Horizon 2020 Framework Programme of the European Union

....
....
....

12. What gender do you identify as? *

- ○ Male
- ○ Female
- ○ Prefer not to say

13. What is your age? *

- ○ 0-15 years old
- ○ 16-30 years old
- ○ 31-45 years old
- ○ 46-60 years old
- ○ 61+

14. What is the highest degree or level of education you have completed? *

In ○middle school
- ○ In high school
- ○ High school degree or equivalent Bachelor's degree (e.g. BA, BS)
- ○ Master's degree (e.g. MA, MS, MEd)
- ○ ○ Professional degree (e.g. MD, DDS, DVM) Doctorate (e.g. PhD, EdD)
- ○ Prefer not to say

15. Please check the box to confirm that you filled out the survey and are eligible for $10 Amazon gift card *

- ○ I confirm

Thank you! The first 100 respondents are eligible to receive a $10 Amazon gift card via email. Please enter your email address below so we can send you the card if you match! To protect your anonymity, we cannot make any connection between your email and your survey.

....

Co-funded by the Horizon 2020 Framework Programme of the European Union