Intelligent Recommendations for Citizen Science

Na'ama Dayan
Ben-Gurion University, Israel
namadayan@gmail.com

Kobi Gal
Ben-Gurion University, Israel
University of Edinburgh, U.K.
kobig@bgu.ac.il

Avi Segal
Ben-Gurion University, Israel
avisegal@gmail.com

Guy Shani
Ben-Gurion University, Israel
shanigu@bgu.ac.il

Darlene Cavalier
Arizona State University, USA
darlene@scistarter.com

ABSTRACT

Citizen science refers to scientific research that is carried out by volunteers, often in collaboration with professional scientists. The spread of the internet has significantly increased the number of citizen science projects and allowed volunteers to contribute to these projects in dramatically new ways. For example, SciStarter, our partners in the project, is an online portal that offers more than 3,000 affiliate projects and recruits volunteers through media and other organizations, bringing citizen science to people. Given the sheer size of available projects, finding the right project, which best suits the user preferences and capabilities, has become a major challenge and is essential for keeping volunteers motivated and active contributors. This paper addresses this challenge by developing a system for personalizing project recommendations in the SciStarter ecosystem. We adapted several recommendation algorithms from the literature based on collaborative filtering and matrix factorization. The algorithms were trained on historical data of users' interactions in SciStarter as well as their contributions to different projects. The trained algorithms were deployed in SciStarter in a study involving hundreds of users who were provided with personalized recommendations for projects they had not contributed to before. Volunteers were randomly divided into different cohorts, which varied the recommendation algorithm that was used to generate suggested projects. The results show that using the new recommendation system led people to contribute to new projects that they had never tried before and led to increased participation in SciStarter projects when compared to cohort groups that were recommended the most popular projects, or did not receive recommendations, In particular, the cohort of volunteers receiving recommendations created by an SVD algorithm (matrix factorization) exhibited the highest levels of contributions to new projects, when compared to the other cohorts. A follow-up survey conducted with the SciStarter community confirms that users were satisfied with the recommendation tool and claimed that the recommendations matched their personal interests and goals. Based on the positive results, our recommendation system is now fully integrated with SciStarter. The research has transformed how SciStarter helps projects recruit and support participants and better respond to their needs.

Proceedings of the ImpactRS Workshop at ACM RecSys '20, September 25, 2020, Virtual Event, Brazil.

INTRODUCTION

Citizen science engages people in scientific research by collecting, categorizing, transcribing, or analyzing scientific data [3, 4, 10]. These platforms offer thousands of different projects which advance scientific knowledge all around the world. Through citizen science, people share and contribute to data monitoring and collection programs. Usually this participation is done as an unpaid volunteer. Collaboration in citizen science involves scientists and researchers working with the public. Community-based groups may generate ideas and engage with scientists for advice, leadership, and program coordination. Interested volunteers, amateur scientists, students, and educators may network and promote new ideas to advance our understanding of the world. Scientists can create a citizen-science program to capture more or more widely spread data without spending additional funding. Citizen-science projects may include wildlife-monitoring programs, online databases, visualization and sharing technologies, or other community efforts.

For example, the citizen science portal SciStarter (scistarter.com), which also comprises our empirical methodology, includes over 3,000 projects, and recruits volunteers through media and other organizations (Discover, the Girl Scouts, etc). As of July, 2020, there are 82,014 registered users in SciStarter. Examples of popular projects on SciStarter include iNaturalist [1] in which users map and share observations of biodiversity across the globe; CoCoRaHS[2], where volunteers share daily readings of precipitation; and StallCatchers [3], where volunteers identify vessels in the brain as flowing or stalled. Projects can be taken either online or at a specific physical region. Users visit SciStarter in order to discover new projects to participate in and keep up to date with the community events. Figure 1 shows the User Interface of SciStarter.

According to a report from the National Academies of Sciences, Engineering, and Medicine [19], citizen scientists' motivations are "strongly affected by personal interests," and participants who engage in citizen science over a long period of time "have successive opportunities to broaden and deepen their involvement." Thus, sustained engagement through the use of intelligent recommendations can improve data quality and scientific outcomes for the projects and the public.

Yet, finding the RIGHT project–one that matches interests and capabilities, is like searching for a needle in a haystack [5, 24]. Ponciano et al. [22] who characterized volunteers' task execution

[1]https://scistarter.org/seek-by-inaturalist
[2]https://scistarter.org/cocorahs-rain-hail-snow-network
[3]https://scistarter.org/stall-catchers-by-eyesonalz

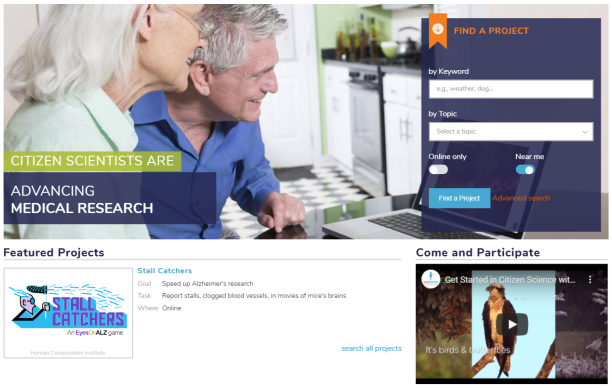

Figure 1: SciStarter User Interface

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. This result is also reflected in users' participation patterns in Scistarter. Figure 2 shows a histogram of the number of projects that users contribute to on the site between 2017 and 2019. As shown by the Figure, the majority of active users in the SciStarter portal do not contribute to more than a single project.

SciStarter employs a search engine (shown in Figure 3) that uses topics, activities, location and demographics (quantifiable fields) to suggest project recommendations. However, recommending projects based on this tool has not been successful. To begin with, our analysis shows that about 80% of users do not use the search tool. Second, those who use the search tool For example, when querying outdoor projects, the search engine recommends the CoCoRaHS project and Globe at Night, in which volunteers measure and submit their night sky brightness observations. But data shows that people who join CoCoRaHS are more likely to join Stall Catchers, an indoor, online project to accelerate Alzheimer's research.

We address this challenge by using recommendation algorithms to match individual volunteers with new projects based on the past history of their interactions on the site [2, 7]. Recommendation systems have been used in other domains, such as e-commerce, news, and social media [8, 13]. However, the nature of interaction in citizen Science is fundamentally different than these domains in that volunteers are actively encouraged to contribute their time and effort to solve scientific problems. Compared to clicking on an advertisement or a product, as is the case for e-commerce and news sites, considerable more effort is required from a citizen science volunteer. Our hypothesis was that personalizing recommendations to users will increase their engagement in the SciStarter portal as measured by the number of projects that they contribute to following the recommendations, and the extent of their contributions.

We attempted to enhance participant engagement to SciStarter projects by matching users with new projects based on past history of their interactions on the site. We adopted 4 different recommendation algorithms to the citizen science domain. The input to the algorithms consists of data representing users' interactions with affiliated projects (e.g., joining or contributing to a project), and users' interactions on the SciStarter portal, (e.g., searching for a project). The output of the algorithm is a function from user profile and past history of interactions on SciStarter to a ranking of 10 projects in order of inferred relevance for the user.

We measured two types of user interactions, which were taken as the input to the algorithms: (1) Interactions with projects: data generated as a result of users' activities with projects, e.g joining a project, making a contribution to a project or participating in a project. (2) Interactions on Scistarter portal, such as searching for a project, or filling a form about the project. The algorithm matches a user profile and his past history of interactions and outputs a ranking of 10 projects in decreasing order of relevance for each user.

We conducted a randomized controlled study, in which hundreds of registered SciStarter users were assigned recommendations by algorithms using different approaches to recommend projects. The first approach personalized projects to participants by using collaborative filtering algorithms (item-based and user-based), and matrix factorization (SVD) algorithms. These algorithms were compared to two non-personalized algorithms: the first algorithm recommended the most popular projects at that point in time, and the second algorithm recommended three projects that were manually determined by the SciStarter admins and custom to change during the study. The results show that people receiving the personalized recommendations were more likely to contribute to new projects that they had never tried before and participated more often in these projects when compared to participants who received non-personalized recommendations, or did not receive recommendations, In particular, the cohort of participants receiving recommendations created by the SVD algorithm (matrix factorization) exhibited the highest levels of contributions to new projects, when compared to the other personalized groups. A follow-up survey conducted with the SciStarter community confirms that the Based on the positive results, our recommendation system is now fully integrated with SciStarter. This research develops a recommendation system for citizen science domain. It is the first study using AI based recommendation algorithms in large scale citizen science platforms.

1 RELATED WORK

This research relates to past work in using AI to increase participants' motivation in citizen science research as well as work in aplying recommendation systems in real world settings. We list relevant work in each of these two areas.

1.1 Citizen Science - Motivation and level of engagement

Online participation in citizen science projects has become very common [21]. Yet, most of the contributions rely on a very small proportion of participants [25]. In SciStarter, the group of participants who contribute to more than 10 projects is less than 10% of all users. 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] showed that participants in citizen science projects are motivated by personal

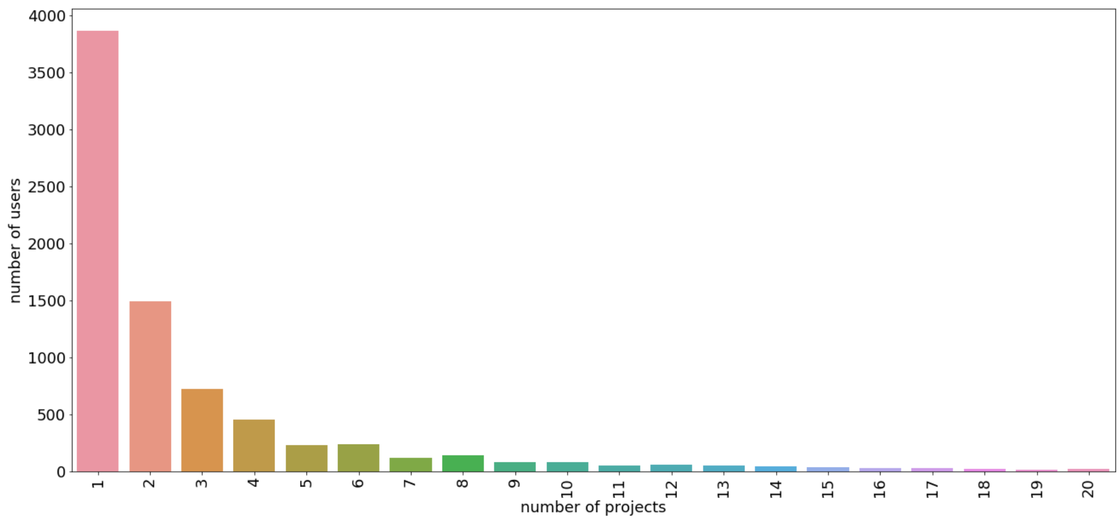

Figure 2: Distribution of user participation in SciStarter projects

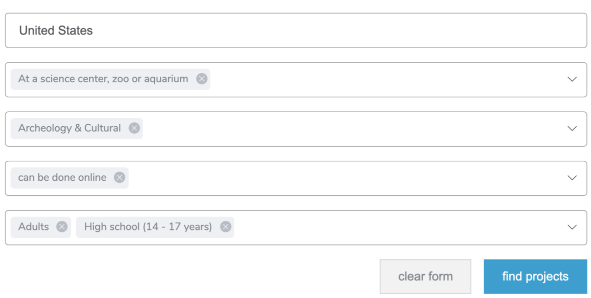

Figure 3: Screenshot of existing search tool showing various criteria

interest and desire to learn something new, as well as by the desire to volunteer and contribute to science. A prior work of Raddic et al. [23] also showed that participants engagement has mainly originated in pure interest in the project topic, such as astronomy and zoology. Yet, as we tested this finding in our collected data, we noticed that user interest is very diverse and does not include only one major topic of interest. Nov et al. [21] explored the different motivations of users to contribute, by separating this question to 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 while quality of contribution is determined by understanding the importance of the task and by the user's reputation. In our work we aimed to increase only the quantity of contributions, since data about the quality of contribution is not available for us.

A significant prior work was done in order to increase participants engagement, which takes into consideration user motivation as well. Segal et al. [29] 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 can be enhanced through the presence of virtual peers.

Ponciano et al. [22] 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 showed 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 in SciStarter's platform instead of in the projects' sites directly, like we do in our research.

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

1.2 Increasing user engagement with recommendations

Similar to our work, other researchers, also tried to increase user engagement and participation by personalized recommendations. Labarthe et al. [16] built a recommender system for students in MOOCs that recommends relevant and rich-potential contacts with other students, based on user profile and activities. They showed 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. [7] 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. This paper showed that users who interacted with the recommendation system increased their chance to finish the MOOC by 270%, compared to users who did not interact with the recommendation system.

Some other studies that concern user engagement with recommendations systems showed how early intervention significantly increase user engagement. Freyne et al. [9] showed that users who received early recommendations in social networks are more likely to continue returning to the site. They showed a clear difference in retention rate between the control group, which has lost 42% of the users, and a group that interacted with the recommendations, which has lost only 24% of the users.

Wu et al. [32], showed how tracking user's clicks and return behaviour in news portals succeeds to increase user engagement with their recommendation system. They 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.

Lin et al. [18], developed a recommendation system for crowdsourcing which incorporates negative implicit feedback into a predictive matrix factorization model. They showed that their models, which consider negative feedback, produce better recommendations than the original MF approach of implicit feedback. They evaluated their findings via experiment with data from Microsoft's internal Universal Human Relevance System and showed that the quality of task recommendations is improved with their models. In our work, we use only positive implicit feedback, due to the low users traffic, where a significant evidence of negative feedback is hard to be found.

Recommendation algorithms are mostly evaluated by their accuracy. The underlying assumption is that accuracy will increase user satisfaction and ultimately lead to higher engagement and retention rate. However, past research has suggested that accuracy does not necessarily lead to satisfaction. Wu et al [31] investigated the effects of popular approaches such as collaborative-filtering and content-based to see if they have different effects on user satisfaction. Results of the study suggested that product awareness (the set of products that the user is initially aware of before using any recommender system) plays an important role in moderating the impact of recommenders. Particularly, if a consumer had a relatively niche awareness set, chances are that content based systems would garner more positive responses on the satisfaction of the user. On the other hand, they showed that users who are more aware of popular items, should be targeted with collaborative filtering systems instead. A subsequent work of Nguyen et al [20], showed that individual users' preferences for the level of diversity, popularity, and serendipity in recommendation lists cannot be inferred from their ratings alone. The paper suggested that user satisfaction can be improved by integrating users' personality traits into the process of generating recommendations, which were obtained by a user study.

2 METHODOLOGY

Our goals for the research project were to (1) help users discover new projects in the SciStarter ecosystem - matching them with projects that are suitable to their preferences. (2) learn user behavior in SciStarter, and develop a recommendation system which will help increase the number of project they contribute to. (3) measure users' satisfaction with the recommendation system.

We adopted several canonical algorithms from the recommendation systems literature: CF user based [28], CF item based [28], Matrix Factorization [27], Popularity [1]. These approaches were chosen as they are all based on analyzing the interactions between users and items and do not rely on domain knowledge which is lacking (such as project's location, needed materials, ideal age group etc.). Each algorithm receives as input a target user and the number of recommendations to generate (N). The algorithm returns a ranking of top N projects in decreasing order of relevance for the user. We provide additional details about each algorithm below.

2.0.1 User-based Collaborative Filtering. In this algorithm, the recommendation is based on user similarities [28]. The ranking of a project for a target user is computed by comparing users who interacted with similar projects. We use a KNN algorithm [6] to find similar users, where the similarity score for user vector U1 and user vector U2 from the input matrix, is calculated with cosine similarity.

$$Similarity(U1, U2) = \frac{U1 * U2}{||U1||||U2||}$$

We chose the value of K to be the minimal number such that the number of new projects in the neighborhood of similar users to the target user equaled the number of recommendations. In practice K was initially chosen to be 100 and increased until this threshold was met. This was done so that there will always be sufficient number of projects to recommend for users.

2.0.2 Item-based Collaborative Filtering. In this algorithm the recommendation is based on project similarity [28]. The algorithm generates recommendations based on the similarity between projects calculated using people's interaction with these projects. Similarity score for project vector P1 and project vector P2 from the input matrix, is calculated with cosine similarity.

$$Similarity(P1, P2) = cos(\theta) = \frac{P1 * P2}{||P1||||P2||}$$

The algorithm then recommends on the top-N most similar projects to the set of projects the user has interacted with in the past.

2.0.3 Matrix Factorization - SVD. The Matrix factorization algorithm (SVD) directly predicts the relevance of a new project to a target user by modeling the user-project relationship [14, 27]. This model-based algorithm (as opposed to the two memory based algorithms presented earlier) was chosen since it is one of the leading recommendation system algorithms [11, 14, 26]. SVD uses a matrix where the users are rows, projects are columns, and the entries are values that represent the relevance of the projects to the users. This users-projects matrix is often very sparse and has many missing values, since users engage with a very small portion of all the available items.

The algorithm estimates the relevance of a target project for a user by maintaining a user model and a project model that include hidden variables (latent factors) that can affect how users choose items. These variables have no semantics, they are simply numbers in a matrix; in reality, aspects like gender, culture, age etc. may affect the relevance, but we do not have access to them.

The singular value decomposition (SVD) of any matrix R is a factorization of the form USV^T. This algorithm is used in recommendation systems 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. As mentioned above, the matrix R includes missing values as users did not participate in all projects. We estimate the missing values which reflect how satisfied will the user be with 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 rectangular diagonal matrix with non-negative real numbers on the diagonal, while V^T is a right singular matrix, indicating the similarity between items and latent factors. 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, and compare between their vectors' representations. Let $\hat{R}$ be the estimation of the original matrix R. Given $\hat{R}$, which includes predictions for all the missing values in R, we can rank each project for a user, by its score in $\hat{R}$. The projects with the highest ranking are then recommended to the user. In our settings, like in the other algorithms described before, $\hat{R}$ is a binary matrix.

3 RESULTS

The first part of the study compares the performance of the different algorithms on historical SciStarter Data. The second part of the study implements the algorithms in the wild, and actively assigns recommendations to users using the different algorithms.

Of the 3000 existing projects SciStarter offers, 153 projects are affiliate projects. 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. As data of contributions and participation only existed for the affiliate projects, we only used these projects in the study.

3.1 Offline Study

The training set for all algorithms consisted of data collected between January 2012 to September 2019. It included 6353 users who contributed to 127 different projects. For the collaborative filtering and SVD algorithm, we restricted the training set to users that made at least two activities during that time frame, whether contributing to a project or interacting on the SciStarter portal. 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. As a baseline, we also considered an algorithm that recommends project according to decreasing order of popularity, measured by the number of users who contribute to the project [1].

We evaluate the top-n recommendation result using precision and recall metrics with varying number of recommendations.

Fig 4 shows results of the precision and recall curves for the 4 examined algorithms. As can be seen from the figure, user-based collaborative filtering and SVD are the best algorithms and their performance is higher than Popularity and Item based collaborative filtering. The Popularity recommendation algorithm generated the lowest performance.

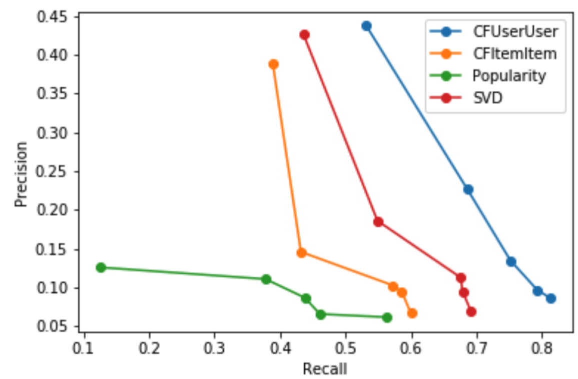

Figure 4: Precision/Recall results on offline data

3.2 Online Study

The second part of the study was an online experiment. Users who logged on to SciStarter starting on December 2nd, 2019 were randomly assigned to one of 5 cohorts, each providing recommendations based on different algorithm: (1) Item-based Collaborative Filtering, (2) User-based Collaborative Filtering, (3) Matrix Factorization, (4) Most popular projects, (5) Promoted projects. Promoted projects were manually determined by SciStarter and often aligned with social initiatives and current events. Among these projects are GLOBE Observer Clouds[4], Stream Selfie [5] and TreeSnap [6]. Another example is FluNearYou, in which individuals report flu symptoms online, was one of the promoted projects during the COVID-19 outbreak. These projects change periodically by the SciStarter administrators.

The recommendation tool was active on SciStarter for 3 months. Users who logged on during that time were randomly divided into cohorts, each receiving a recommendation from a different algorithm. Each cohort had 42 or 43 users. The recommendations were embedded in the user's dashboard in decreasing order of relevance, in sets of three, from left to write. Users could scroll to reveal more projects in decreasing or increasing order of relevance. Figure 5 shows the top three recommended projects for a target user.

All registered users in SciStarter received notification via email about the study, stating that the "new SciStarter AI feature provides personalized recommend projects based on your activity and interests." A link to a blog post containing more detailed explanations of recommendation algorithms, their role in the study, emphasizing that' 'all data collected and analyzed during this experiment on SciStarter will be anonymized." Also, users are allowed to opt

[4]https://scistarter.org/globe-observer-clouds
[5]https://scistarter.org/stream-selfie
[6]https://scistarter.org/treesnap

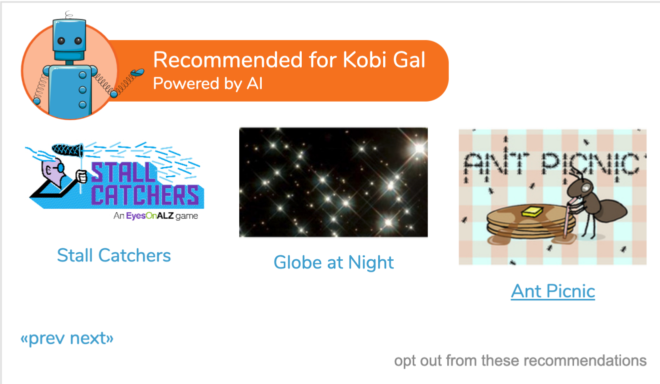

Figure 5: Screenshot of recommendation tool

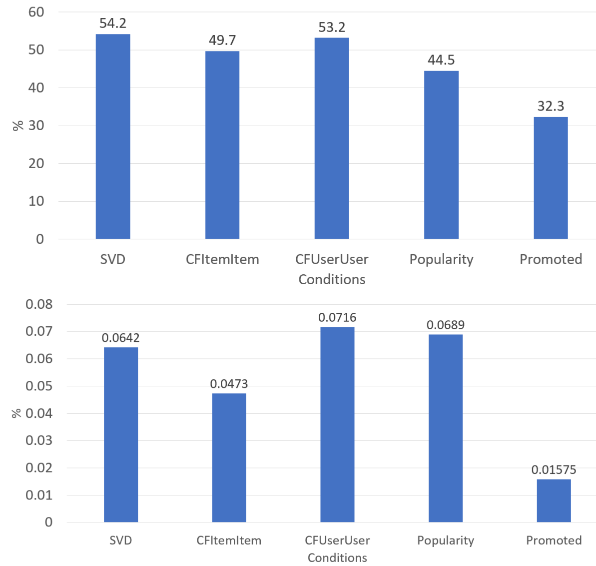

Figure 6: Click through rate (top) and Hit rate (bottom) measures for online study

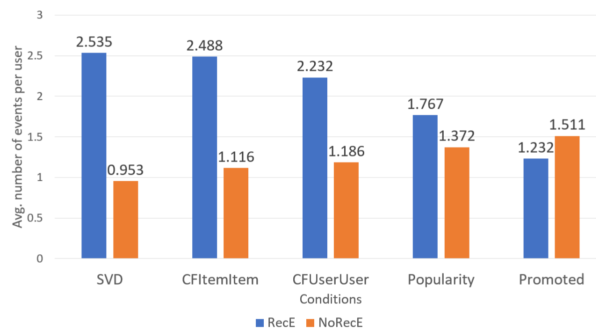

Figure 7: Average activities on recommended projects (RecE), and on non-recommended projects (NoRecE) for each condition

out of receiving recommendations at any point, by clicking on the link “opt out from these recommendations" in the recommendation panel. In practice, none of the participants selected the opt out option at any point in time.

Figure 6 (top) shows the average click trough rate (defined as the ratio recommended projects that the users accessed) and Figure 6 (bottom) shows the average hit rate (defined as the percentage of instances in which users accessed at least one project that was recommended to them). As shown by the Figure, both measures show a consistent trend, in which the user-based collaborative algorithms achieved the best performance, while the baseline method achieving worse performance. Despite the trend, the differences between conditions were not statistically significant in the $p < 0.05$ range. We attribute this to the fact that we measured clicks on recommended projects rather than actual contributions which is the most important aspect for citizen science.

To address this gap we defined two new measures that consider the contributions made by participants to projects, which considers the system utility and identified by Gunawardana and Shani [12]. The measures include the average number of activities that users carried out in recommended projects (RecE), and the average number of activities that users carried out in non-recommended projects (NoRecE). Figure 7 compares the different algorithms according to these two measures. The results show that users assigned to the intelligent recommendation conditions performed significantly more activities in recommended projects than those assigned to the Popularity and Baseline conditions. Also, users in the SVD algorithm performed significantly less activities in non-recommended projects than the Popularity and Baseline conditions. These results were statistically significant according to Mann-Whitney tests (see Appendix for details).

Lastly, we measure the average number of sessions for users in the different conditions, where sessions are defined as a continuous length of time in which the user is active in a project. Figure 8 shows the average number of sessions for users in the different cohorts, including the number of sessions for the historical data used to train the algorithms, in which no recommendations were provided. The results show that users receiving recommendations from the personalized algorithms performed more sessions than the number of sessions in historical data. These results are statistically significant. Although there is a clear trend that users in the *SVD* condition achieved the highest number of sessions, these results were not significant in the $p < 0.05$ range.

To explain the success of SVD’s good performance in the online study, we note first that SVD is considered as a leading algorithm in the domain of recommendation systems [11, 26]. Second, in our setting SVD tended to generate recommendations that participants had not heard about before, which the survey reveals to be more interesting to them. One participant remarked: "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".

Lastly, we note the obstacles we encountered when carrying out the study. The first obstacle we encountered was the small number of relevant projects that could be recommended. Out of

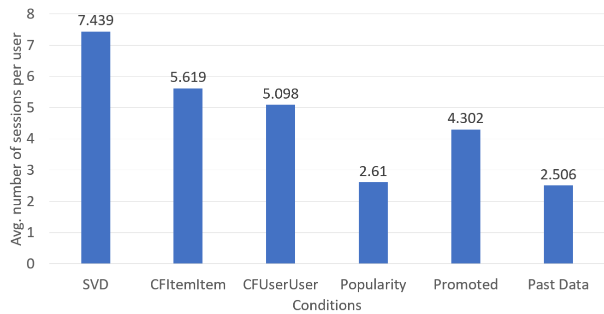

Figure 8: Average number of sessions for each condition

almost 3000 projects that SciStarter offers, we restricted ourselves to about 120 projects are affiliate projects which actively provide data of users' interactions. Another obstacle was that we were constrained to a subset of users who log on to SciStarter and use it as a portal to contributing to the project, rather than accessing the project directly. Out of the 65,000 registered users of SciStarter, only a small percentage are logged in to both SciStarter and an affiliate project. As a result, we have relatively few users getting recommendations. In addition, some of SciStarter's projects are location-specific and can only be done by users in the same physical location. (e.g collecting a water sample from a particular lake located in a particular city). Therefore, we kept track of users' location and restricted our recommendation system to be a location-based system, which recommends users with projects they are able to participate in.

3.3 User Study

In order to learn what is the users' opinion on the recommendations, and their level of satisfaction, we conducted a survey with SciStarter's users. Our survey was sent to all SciStarter community users. 138 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. The majority of users (75%) were very satisfied with the recommendation tool and claimed that the recommendations matched their personal interests and goals. The majority of users (54%) reported they have clicked on the recommendations and visited the project's site, while only 8% of users did not click the recommendation or visited the project site. Interestingly, 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 [30], 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".

The survey provided evidence for the positive impact of using the recommendation systems in SciStarter, which include the following comments. "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", and "I like how personalized recommendations are made for citizen science users".

4 CONCLUSION AND FUTURE WORK

This work reports on the use of recommendation algorithms to increase engagement of volunteers in citizen science, in which volunteers collaborate with researchers to perform scientific tasks. These recommendation algorithms were deployed in SciStarter, a portal with thousands of citizen science projects, and were evaluated in an online study involving hundreds of users who were informed about participating in a study involving AI based recommendation of new projects. We trained different recommendation algorithms using a combination of data including users' behavior in SciStarter as well as their contributions to the specific project. Our results show that using the new recommendation system led people to contribute to new projects that they had never tried before and led to increased participation in SciStarter projects when compared to a baseline cohort group that did not receive recommendations. The outcome of this research project is the AI-powered Recommendation Widget which has been fully deployed in SciStarter. This project has transformed how SciStarter helps projects recruit and support participants and better respond to their needs. It was so successful in increasing engagement, that SciStarter has decided to make the widget a permanent feature of their site. This will help support deeper, sustained engagement to increase the collective intelligence capacity of projects and generate improved scientific, learning, and other outcomes. The results of this research have been featured on the DiscoverMagazine.com [7]. While we observed significant engagement with the recommendation tool, one may consider adding explanations to the recommendations in order to increase the system's reliability and user's satisfaction with it. Moreover, we plan to extend the recommendation system to include content based algorithms, and test its performance as compared to the existing algorithms. We believe that integrating content in Citizen Science domain can be very beneficial. Even though users tend to participate in a variety of different projects, we want to be able to capture more intrinsic characteristic of the projects, such as the type of the task a user has to perform or the required effort.

REFERENCES

[1] Hyung Jun Ahn. 2006. Utilizing popularity characteristics for product recommendation. *International Journal of Electronic Commerce* 11, 2 (2006), 59–80.
[2] Xavier Amatriain. 2013. Big & personal: data and models behind netflix recommendations. In *Proceedings of the 2nd international workshop on big data, streams and heterogeneous source Mining: Algorithms, systems, programming models and applications*. ACM, 1–6.
[3] Rick Bonney, Caren B Cooper, Janis Dickinson, Steve Kelling, Tina Phillips, Kenneth V Rosenberg, and Jennifer Shirk. 2009. Citizen science: a developing tool for expanding science knowledge and scientific literacy. *BioScience* 59, 11 (2009), 977–984.

[7] https://www.discovermagazine.com/technology/ai-powered-smart-project-recommendations-on-scistarter

[4] Dominique Brossard, Bruce Lewenstein, and Rick Bonney. 2005. Scientific knowledge and attitude change: The impact of a citizen science project. *International Journal of Science Education* 27, 9 (2005), 1099–1121.
[5] Hillary K Burgess, LB DeBey, HE Froehlich, Natalaie Schmidt, Elli J Theobald, Ailene K Ettinger, Janneke HilleRisLambers, Joshua Tewksbury, and Julia K Parrish. 2017. The science of citizen science: Exploring barriers to use as a primary research tool. *Biological Conservation* 208 (2017), 113–120.
[6] Sahibsingh A Dudani. 1976. The distance-weighted k-nearest-neighbor rule. *IEEE Transactions on Systems, Man, and Cybernetics* 4 (1976), 325–327.
[7] Surabhi Dwivedi and VS Kumari Roshni. 2017. Recommender system for big data in education. In *2017 5th National Conference on E-Learning & E-Learning Technologies (ELELTECH)*. IEEE, 1–4.
[8] Daniel M Fleder and Kartik Hosanagar. 2007. Recommender systems and their impact on sales diversity. In *Proceedings of the 8th ACM conference on Electronic commerce*. ACM, 192–199.
[9] Jill Freyne, Michal Jacovi, Ido Guy, and Werner Geyer. 2009. Increasing engagement through early recommender intervention. In *Proceedings of the third ACM conference on Recommender systems*. ACM, 85–92.
[10] Cary Funk, Jeffrey Gottfried, and Amy Mitchell. 2017. Science news and information today. *Pew Research Center* (2017).
[11] Stephen Gower. 2014. Netflix prize and SVD.
[12] Asela Gunawardana and Guy Shani. 2009. A survey of accuracy evaluation metrics of recommendation tasks. *Journal of Machine Learning Research* 10, 12 (2009).
[13] J Itmazi and M Gea. 2006. The recommendation systems: Types, domains and the ability usage in learning management system. In *Proceedings of the International Arab Conference on Information Technology (ACIT'2006), Yarmouk University, Jordan*.
[14] Yehuda Koren, Robert Bell, and Chris Volinsky. 2009. Matrix factorization techniques for recommender systems. *Computer* 42, 8 (2009), 30–37.
[15] Gitte Kragh. 2016. The motivations of volunteers in citizen science. *environmental SCIENTIST* 25, 2 (2016), 32–35.
[16] Hugues Labarthe, François Bouchet, Rémi Bachelet, and Kalina Yacef. 2016. Does a Peer Recommender Foster Students' Engagement in MOOCs?. *International Educational Data Mining Society* (2016).
[17] Jeffrey Laut, Francesco Cappa, Oded Nov, and Maurizio Porfiri. 2017. Increasing citizen science contribution using a virtual peer. *Journal of the Association for Information Science and Technology* 68, 3 (2017), 583–593.
[18] Christopher H Lin, Ece Kamar, and Eric Horvitz. 2014. Signals in the silence: Models of implicit feedback in a recommendation system for crowdsourcing. In *Twenty-Eighth AAAI Conference on Artificial Intelligence*.
[19] Engineering National Academies of Sciences, Medicine, et al. 2018. *Learning through citizen science: enhancing opportunities by design*. National Academies Press.
[20] Tien T Nguyen, F Maxwell Harper, Loren Terveen, and Joseph A Konstan. 2018. User personality and user satisfaction with recommender systems. *Information Systems Frontiers* 20, 6 (2018), 1173–1189.
[21] Oded Nov, Ofer Arazy, and David Anderson. 2014. Scientists@ Home: what drives the quantity and quality of online citizen science participation? *PloS one* 9, 4 (2014), e90375.
[22] Lesandro Ponciano and Thiago Emmanuel Pereira. 2019. 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*. ACM, 16.
[23] M Jordan Raddick, Georgia Bracey, Pamela L Gay, Chris J Lintott, Phil Murray, Kevin Schawinski, Alexander S Szalay, and Jan Vandenberg. 2009. Galaxy zoo: Exploring the motivations of citizen science volunteers. *arXiv preprint arXiv:0909.2925* (2009).
[24] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2015. Recommender systems: introduction and challenges. In *Recommender systems handbook*. Springer, 1–34.
[25] Dana Rotman, Jenny Preece, Jen Hammock, Kezee Procita, Derek Hansen, Cynthia Parr, Darcy Lewis, and David Jacobs. 2012. Dynamic changes in motivation in collaborative citizen-science projects. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*. 217–226.
[26] Rowayda A Sadek. 2012. SVD based image processing applications: state of the art, contributions and research challenges. *arXiv preprint arXiv:1211.7102* (2012).
[27] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. 2002. Incremental singular value decomposition algorithms for highly scalable recommender systems. In *Fifth international conference on computer and information science*, Vol. 1. Citeseer.
[28] J Ben Schafer, Dan Frankowski, Jon Herlocker, and Shilad Sen. 2007. Collaborative filtering recommender systems. In *The adaptive web*. Springer, 291–324.
[29] Avi Segal, Kobi Gal, Ece Kamar, Eric Horvitz, and Grant Miller. 2018. Optimizing Interventions via Offline Policy Evaluation: Studies in Citizen Science. In *Thirty-Second AAAI Conference on Artificial Intelligence*.
[30] Avi Segal, Ya'akov Gal, Robert J Simpson, Victoria Victoria Homsy, Mark Hartswood, Kevin R Page, and Marina Jirotka. 2015. Improving productivity in citizen science through controlled intervention. In *Proceedings of the 24th International Conference on World Wide Web*. 331–337.
[31] Ling-Ling Wu, Yuh-Jzer Joung, and Jonglin Lee. 2013. Recommendation systems and consumer satisfaction online: moderating effects of consumer product awareness. In *2013 46th Hawaii International Conference on System Sciences*. IEEE, 2753–2762.
[32] Qingyun Wu, Hongning Wang, Liangjie Hong, and Yue Shi. 2017. Returning is believing: Optimizing long-term user engagement in recommender systems. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. ACM, 1927–1936.

A APPENDIX

A.1 Significance tests - number of activities

A Mann-Whitney test was conducted to compare between each condition in the online experiment. Table 1 presents the results of the pairwise tests for the measures RecE and NoRecE that are significant.

Condition1	Condition2	U	n1	n1	DV	Significant
CFUserUser	Popularity	473.5	43	43	RecE	Yes
CFUserUser	Baseline	406.0	43	43	RecE	Yes
CFItemItem	Popularity	458.5	43	43	RecE	Yes
CFItemItem	Baseline	396.0	43	43	RecE	Yes
SVD	Popularity	433.0	42	43	RecE	Yes
SVD	Baseline	371.5	42	43	RecE	Yes
SVD	CFItemItem	731.0	42	43	RecE	Yes
SVD	Baseline	729.0	42	43	NoRecE	Yes

Table 1: Online Metrics - Mann Whitney significance test with p<0.05. DV=Dependent Variable

A.2 Significance tests - number of sessions

A Mann-Whitney test was conducted to compare between each condition in the online experiment, including the historical data used to train the algorithms, called past-data. Table 2 presents the results of the pairwise tests that are significant.

Condition1	Condition2	U	n1	n2	Significant
CFUserUser	Past-Data	5898.0	43	557	Yes
CFItemItem	Past-Data	6502.0	43	557	Yes
SVD	Past-Data	7284.0	42	557	Yes
Popularity	Past-Data	6683.5	43	557	Yes
Baseline	Past-Data	6978.5	43	557	Yes

Table 2: Number of sessions in SciStarter - Mann Whitney significance test with p<0.05