Bumblebee Friendly Planting Recommendations with Citizen Science Data

Agung Toto Wibowo
Computing Science
University of Aberdeen
wibowo.agung@abdn.ac.uk

Advaith Siddharthan
Knowledge Media Institute
The Open University
advaith.siddharthan@open.ac.uk

Helen Anderson
School of Biological Sciences
University of Aberdeen
helen.anderson1@abdn.ac.uk

Annie Robinson
School of Biological Sciences
University of Aberdeen
annierobinson@abdn.ac.uk

Nirwan Sharma
Computing Science
University of Aberdeen
n.sharma@abdn.ac.uk

Helen Bostock
Royal Horticultural Society
helenbostock@rhs.org.uk

Andrew Salisbury
Royal Horticultural Society
andrewsalisbury@rhs.org.uk

Richard Comont
Bumblebee Conservation Trust
richard.comont@
bumblebeeconservation.org

René van der Wal
School of Biological Sciences
University of Aberdeen
r.vanderwal@abdn.ac.uk

ABSTRACT

Several citizen science projects engage with the public around pollinator species, typically requesting data (e.g. in the form of photo-records of different species tagged by place and date). While such projects help scientists collect data, these data are rarely fed back to the public in any meaningful manner. In this paper, we address this through a recommender system based on Matrix Factorization over a matrix of observed bumblebee–plant interactions derived from data submitted to a citizen science project BeeWatch. The system recommends pollinator-friendly plants for domestic gardens and takes into account both the fact that different bumblebee species exhibit differing preferences for flowers, and that plants flower at different times of the year. The goal is to attract a range of bumblebee species to a garden and to ensure that these species have sufficient food sources through the season.

CCS CONCEPTS

• **Human-centered computing** → *Social recommendation*; • **Applied computing** → *Interactive learning environments*; *Life and medical sciences*;

KEYWORDS

Recommender Systems, Planting Advice, Pollinator Friendly Plants, Bumblebees, Matrix Factorization

ACM Reference format:
Agung Toto Wibowo, Advaith Siddharthan, Helen Anderson, Annie Robinson, Nirwan Sharma, Helen Bostock, Andrew Salisbury, Richard Comont, and René van der Wal. 2017. Bumblebee Friendly Planting Recommendations with Citizen Science Data. In *Proceedings of CitRec, Como, Italy, August 27, 2017,* 6 pages.
https://doi.org/10.1145/3127325.3128330

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CitRec, August 27, 2017, Como, Italy
© 2017 Association for Computing Machinery.
ACM ISBN 978-1-4503-5370-0/17/08…$15.00
https://doi.org/10.1145/3127325.3128330

1 INTRODUCTION

There is much concern among scientists and society about the decline in pollinator populations, which has been linked to several factors including pesticide use and the loss of wildflower habitats. Environmental challenges such as biodiversity loss are increasingly seen as impossible to fully comprehend or solve [6] and it is generally accepted that fundamentally different approaches are needed. This has led to a strong call for "collaborative research efforts among scientists, educators, and the public, linking science and society with place and identity, through more effective processes of public engagement and learning that can result in meaningful socioecological outcomes" [22].

With the growing numbers of people living in urban areas, domestic gardens have today become an important habitat for pollinating insects such as bumblebees [15]. Several citizen science projects today take advantage of this and engage with the public around pollinator species, often soliciting data in the form of photographs taken in domestic gardens and elsewhere [17, 18, 20]. The photographs are then identified at species level, either by other users of the citizen science portal, or by experts overseeing the portal, thus helping scientists collect data at geographical scale to model species distributions. However, these data are rarely fed back to the public in any meaningful manner, resulting in missed opportunities for digital engagements that promote learning, participatory action and behavioural or attitudinal change [16].

To directly address this, we present a recommender system to recommend bumblebee-friendly plants for domestic gardens, taking into account both the fact that different bumblebee species exhibit preferences for different flowers, and that plants flower at different times of the year. The goal is to recommend plants that attract a range of bumblebee species to a garden and to ensure that these species have sufficient food sources all the way through the season.

Crucially, we make these recommendations in the context of data provided by participants in a real world citizen science project, BeeWatch[1].

BeeWatch, a collaboration between the University of Aberdeen and the Bumblebee Conservation Trust[2], invites members of the British public to submit photographs they have taken of bumblebees and to use an online identification key to classify (label) their photograph as one of 22 possible species present in the UK [17]. To ensure data quality, each identification is verified either through a Bayesian consensus model using independent identifications from other BeeWatch participants, or by a bumblebee expert. In order to improve participants' species identification skills and enhance their user experience, BeeWatch also generates automated feedback on their submission using natural language generation (NLG) technology [2, 21]. To further engage with participants around the data they contribute, and in particular to target wider learning, participatory action and attitudinal change, we chose to implement a recommender system that can offer gardening advice. Participants receive recommendations either interactively by engaging with the Planting for Pollinators Tool accessible from the BeeWatch website, or through recommendations embedded within the automatically generated feedback BeeWatch participants receive on submitting a bumblebee record.

Recommender systems have been successfully implemented in several real world applications. E-commerce sites (such as e-Bay, Amazon, Netflix, Spotify, etc.) use recommender systems to boost their cross-selling by providing alternative products to users. Some approaches, such as collaborative filtering (CF) [5], have been developed by researchers [3] to calculate (and model) similarity from user–item interactions [13]. The interactions themselves might be in the form of transactions (such as buying items) or explicit user ratings of products [3, 8, 14]. One popular algorithm for CF is matrix factorization (MF), which tries to model user–item ratings as a set of latent factors [10].

Unlike e-commerce applications, in BeeWatch we do not have traditional transactional or rating data connecting bumblebee species to plant species. Rather, user submitted photographs contain a species of bumblebee, and as photos were typically taken in domestic gardens, the majority of records on BeeWatch have species level identifications by users of the plant as well as the bumblebee. We use these records to create a bumblebee–plant interaction matrix that records the number of times each bumblebee species has been recorded on each plant species. As there are 242 species of plant and 22 species of bumblebee in our BeeWatch database, even with over 10,000 interactions in the database, the interaction matrix is still quite sparse (cf. Table 2). However, there are likely several latent factors in the data that capture similarity between plants and between bumblebees, and we believe MF methods can help overcome this data sparseness.

In this paper we discuss three recommender systems tasks and how they are deployed: (1) bumblebee–plant interaction frequency approximation, (2) year-round planting recommendations, and (3) contextual recommendations to recorders. In the first task, we use MF to approximate missing values to produce a full interaction matrix as output. In the second task, we combine recommendations produced by MF with knowledge about plants in a user's garden, as provided by them, and the flowering times of plants to provide planting recommendations that provide food for the bumblebee species expected in the garden all through the season. In the third task, we provide contextual recommendations to users at the point that they submit records to BeeWatch.

[1]www.abdn.ac.uk/research/beewatch
[2]www.bumblebeeconservation.org

Table 1: List of Mathematical Notations

Notations	Descriptions
$l = 12$	number of months in the year
$m = 22$	number of bumblebee species in dataset
$n = 242$	number of plant species in dataset
b_x	a bumblebee b with species id $x = 1..22$
p_y	a plant p with species id $y = 1..242$
f_{b_x,p_y}	frequency of observed records for bumblebee b_x and plant p_y
$\hat{f}_{b_x,p_y}$	estimated interaction frequency for bumblebee b_x and plant p_y
$V^{m\times n} = (v_{x,y})$	matrix of frequencies of observed interactions, where $v_{x,y} = f_{b_x,p_y}$;
$\hat{V}^{m\times n} = (\hat{v}_{x,y})$	matrix of estimated frequencies of interactions, where $\hat{v}_{x,y} = \hat{f}_{b_x,p_y}$;
$S^{n\times l} = (s_{x,y})$	matrix of flowering status, where $s_{x,y} = 1$ if p_x flowers in month y $s_{x,y} = 0$ if p_x does not flower in month y
$\mathcal{W}^{m\times n\times l} = (w_{x,y,z})$	tensor of interactions by flowering status, where $w_{x,y,z} = \hat{v}_{x,y} * s_{y,z}$

2 RECOMMENDING BUMBLEBEE-FRIENDLY PLANTS

Consider a set of bumblebee species $B = \{b_1, b_2, \cdots, b_m\}$ that can interact with a set of plant species[3] $P = \{p_1, p_2, \cdots, p_n\}$. Since the same bumblebee–plant interaction can be recorded many times (at different times or by different citizen scientists, depending on how common it may be), each interaction is denoted as a frequency f_{b_x,p_y}, the total number of records in BeeWatch for the interaction between a bumblebee species b_x and a plant species p_y. We store these interactions in a bumblebee–plant matrix $V^{m\times n}$ where each cell $v_{x,y}$ contains the interaction frequency f_{b_x,p_y}. Table 1 summarises the notation used in this paper, and Table 2 provides a statistical summary of the BeeWatch dataset for bumblebee–plant interactions.

Note that the density of the matrix is around 20%, i.e. four fifths of possible species level interactions between bumblebees and plants have never been recorded in BeeWatch. This does not mean that they do not occur, just that our database is sparse. As BeeWatch users are not given any direction as to what records to submit, the database reflects a random sample that is naturally skewed in favour of records involving common plant and bumblebee species

[3]This paper will refer throughout to "plant species". However, this is a simplification as the classification of plants used is either at genera or species level depending on the level of detail within the system.

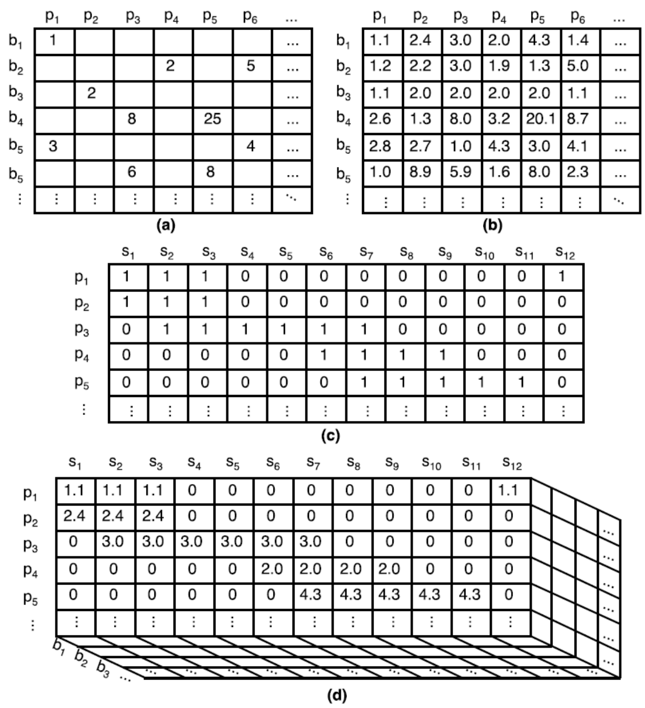

Figure 1: Visual representation of matrices used in this paper and defined in Table 1: (a) Observed interaction frequency matrix V, (b) Estimated interaction frequency matrix $\hat{V}$, (c) Binary matrix of flowering status by month S, (d) Tensor of interactions by flowering status $\mathcal{W}$

(the most commonly recorded interaction has a frequency of 311, and the least commonly recored interactions have a frequency of 1). To approximate the missing interactions frequencies, we used a Matrix Factorization algorithm over the bumblebee–plant interaction matrix $V^{m \times n}$.

There are many approaches to collaborative filtering for estimating missing values in transaction matrices. Some approaches calculate similarity between users or items [7, 19], while other approaches use matrix factorization techniques to decompose the rating matrix into two (or more) matrices. The first winner [9] of the Netflix prize reported that matrix factorization has many benefits for overcoming common problems in recommender systems such as data sparsity and cold start [23].

Matrix factorization (MF) can be defined as producing two factor matrices, say $W = \left[w_{ij}\right] \in \mathbb{R}^{m \times k}$ and $H = \left[h_{ij}\right] \in \mathbb{R}^{k \times n}$ from one known matrix $V = \left[v_{ij}\right] \in \mathbb{R}^{m \times n}$, so the product of W and H is (approximately) equal to V:

$$WH = \hat{V} \approx V \ , \tag{1}$$

where each cell in $\hat{V}$ is computed as:

$$\hat{v}_{xy} \approx \sum_{i=1}^{k} w_{xi} h_{iy} \tag{2}$$

Table 2: Statistics of BeeWatch Interaction Dataset

No. of Bumblebee Species	22
No. of Plant Species	242
No. of unique Bumblebee–Plant Interactions	1,111
Total No. of possible interactions (22 × 242)	5,324
Matrix Density (1,111/5,324)	20.87%
Total No. of Bumblebee–Plant Interaction Records	10,647
Range of observed interaction frequencies	(1–311)

and the choice of k, the number of latent dimensions, is typically determined empirically.

There are many algorithms for MF, such as Multiplicative, Gradient Descent and Alternating Least Square [1, 4, 11, 12]. These algorithms aim to minimize the difference between the known values in matrix V and the corresponding values in its multiplicative form WH (the cost function) through an iterative process. When the factors W and H are computed in this manner, it has been found that the product WH provides values for missing cells in V, and that these turn out to be good estimates of these missing ratings.

2.1 Interaction Frequency Estimation

As described earlier, our first task is to estimate missing values in a bumblebee–plant frequency of observation matrix ($V^{m \times n}$) where the frequencies of observed cells $v_{x,y}$ range from 1 to 311.

2.1.1 Method. Unlike the transaction matrices of fixed scale product ratings commonly used in recommender systems, our bumblebee–plant frequency matrix (V) has values ranging from 1 to 311, and it is clear that these are not equally spaced: A difference of, say, 10 at the lower end of the scale (e.g. 1 and 11) is much more important than at the high end (e.g. 300 and 310). We therefore normalized the frequencies of interaction using $log_2(f_{b_x,p_y})$. After normalisation, the range of values is 0 ($log_2 1$) to 8.28 ($log_2 311$), roughly equivalent to a ten point scale.

We used the iterative gradient descent matrix factorization algorithm with 100 iterations, a commonly used setting for recommendation systems in the rating predictions task. In this study we varied k, the number of latent dimensions in MF, (k = 1...10). We report Root Mean Square Error (RMSE) using 4-fold cross-validation methodology, i.e we partitioned the dataset into 4 parts randomly, and in each fold, we used one part to test while training on the other three parts, reporting average results over all four folds.

2.1.2 Results. The results are reported in Table 3. The first column reports the average RMSE for raw frequencies (which range from 1 to 311). The second column reports RMSE over the normalised data (which range from 0 to 8.28). For estimating raw frequencies, the lowest error rate is obtained for $k = 7$, while for normalised frequencies, fewer latent dimensions are needed to capture the data, with the best performance obtained using $k = 4$.

We compared the MF method to two standard baselines: (a) "Bumblebee Average", which takes the average of the cells with observations in the row for that bumblebee species (Eqn. 3), and (b) "Plant Average", which takes the average of the cells with observations in the column for that plant species (Eqn. 4).

$$Av(b_x) = \frac{\sum_{j=1}^{n} v_{x,j}}{\sum_{j=1}^{n} \begin{cases} 1, & \text{if } v_{x,j} \geq 1 \\ 0, & \text{otherwise} \end{cases}} \quad (3)$$

$$Av(p_y) = \frac{\sum_{i=1}^{m} v_{i,y}}{\sum_{i=1}^{m} \begin{cases} 1, & \text{if } v_{i,y} \geq 1 \\ 0, & \text{otherwise} \end{cases}} \quad (4)$$

MF outperformed both baselines. Among the baselines, the plant proved to be a better indicator than the bumblebee and $Av(p_y)$, the average of the observed frequencies for different bumblebees interacting with the given plant species p_y gave lower error than $Av(b_x)$, the average of the observed frequencies for the different plants interacting with the given bumblebee species b_x.

2.1.3 Deployment. We used the estimated normalised interaction frequencies obtained with $k = 4$ to allow citizen scientists to browse the data they have collectively recorded in two ways: (a) by "bumblebees preferences" and (b) by "plant attractiveness". In the former, citizen scientists could use an interactive interface to select a bumblebee species and explore the plants they favour. Figure 2 shows a screenshot where a user has selected the Tree bumblebee. The length of horizontal bar represent the "Strength of Prediction" obtained from matrix factorization. Similarly, citizen scientists could interactively select a plant species and explore which bumblebee species they might attract.

2.2 Year-round Planting Recommendations

One important consideration when recommending bumblebee-friendly plants is that different plant species flower at different times of the year. As it is critically important that gardens provide suitable flowers for bumblebees throughout the season, our second, and main, task is to provide year-round planting recommendations.

2.2.1 Method. We collected information about the flowering months for plants from the webpages of the Royal Horticultural

Table 3: Average RMSE Performance for estimating raw frequencies and frequencies normalised by log_2 (lowest error highlighted in green)

Algorithm	k	f_{b_x,p_y}	$log_2(f_{b_x,p_y})$
Bumblebee Average		19.95	1.67
Plant Average		18.57	1.65
Matrix Factorization	1	20.17	1.56
	2	19.08	1.58
	3	17.49	1.53
	4	17.18	1.53
	5	16.71	1.59
	6	17.28	1.57
	7	16.70	1.58
	8	16.95	1.62
	9	17.39	1.61
	10	16.91	1.63

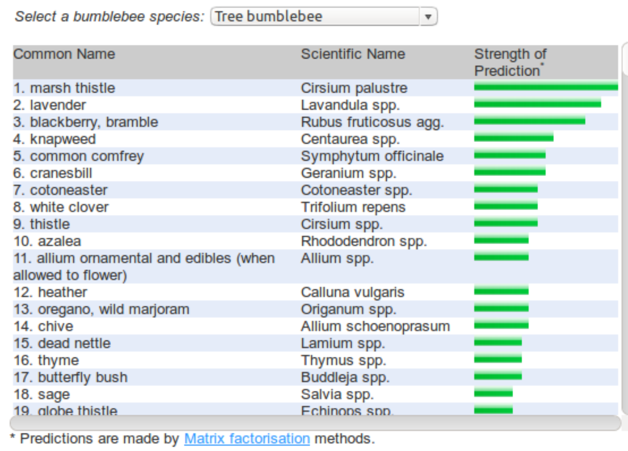

Figure 2: Screenshot of interactive user interface for exploring the plants preferred by different Bumblebee species

Society[4]. Given a set of n plants p_x, we constructed a binary matrix of flowering status $S^{n \times l} = (s_{x,y})$, where $s_{x,y} = 1$ if p_x flowers in month y, $s_{x,y} = 0$ if p_x does not flower in month y. Then, using the estimated bumblebee–plant interactions matrix $\hat{V}^{m \times n}$, we constructed a tensor of interactions by flowering status $\mathcal{W}^{m \times n \times l} = (w_{x,y,z})$, where $w_{x,y,z}$ is the product of the estimated bumblebee–plant interaction frequency $\hat{v}_{x,y}$ and the binary value $s_{y,z}$ from the matrix of flowering status $S^{n \times l}$. The process of constructing this tensor is illustrated in Figure 1. We next narrate how this tensor is used in the user interface.

2.2.2 Deployment. Our interface for year-round planting recommendations first asks the user to select from a menu zero or more plants that they have in their garden. The interface then provides recommendations as shown in the screenshot in Figure 3. In this example, the user has informed us that they have blackberry and lavender in their garden (top section). The interface then lists (middle section) the 10 most likely bumblebees that interact with these plants (using $\hat{V}$). If the user has not selected any plant species, this list defaults to the 10 most frequently recorded bumblebee species. Finally, in the bottom section, the interface makes recommendations of plants that flower at different times through the year and are popular with the 10 bumblebee species identified in the middle section.

The recommendations in the bottom section are constructed as follows. Using the tensor of interactions by flowering status ($\mathcal{W}$) we retrieve the top-N plants for the 10 bumblebee species (middle section) in winter, spring, summer and autumn. We defined winter as December–February, spring as March–May, summer as June–August and autumn as September–November. For example, based on Figure 1(d) we would calculate the weight for plant "p_1" as 3.3 (1.1 + 1.1 + 1.1 in blue column) for winter, 1.1 for spring, 0 for summer and 0 for autumn. Our recommendation is generated by intersecting the top-N list for each season.

[4] www.rhs.org.uk

The list of recommended plants are thus generated to ensure that bumblebees have good availability of flowers through the season.

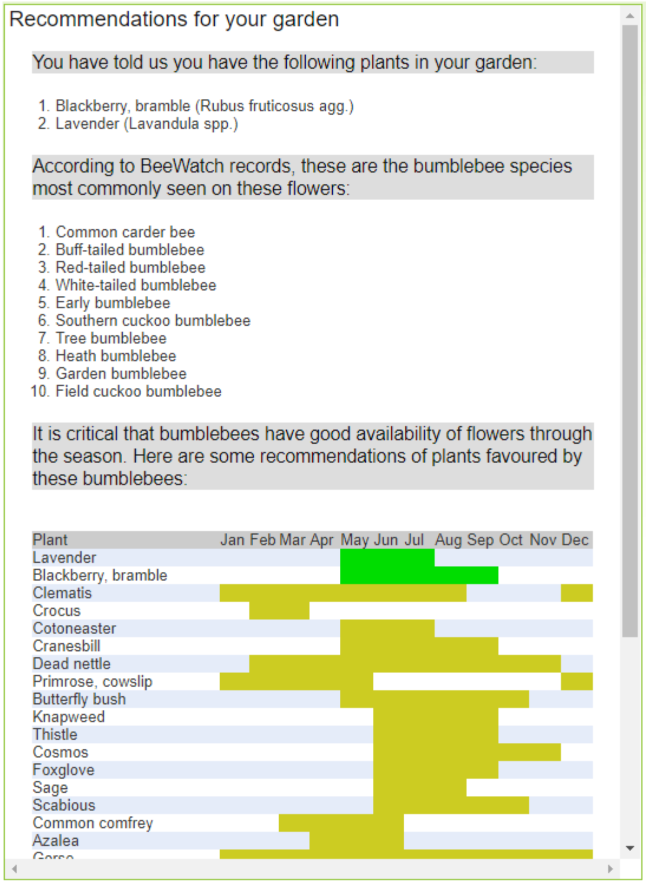

Figure 3: Example of Year-Round Recommendations generated by the interactive user interface

2.3 Contextual Recommendations

In addition to providing web interfaces to recommender systems for citizen scientists to explore the data they collectively submit, we also provide planting recommendations at the time they submit a record to BeeWatch. This is provided by email, in the form of an automatically generated text that takes into account the bumblebee species recorded by the user and the current month. This planting advice (an example is shown in Figure 4) is embedded in a longer formative feedback about the submission (see [2, 21] for details).

3 DISCUSSION AND FUTURE WORK

We have described how we have successfully embedded recommender systems into a citizen science platform. Through this process, we have progressed the platform from being primarily about data collection to one that encourages citizens to think about changes they can make to their local environment to support biodiversity. It is specifically designed to raise awareness of planting issues that impact on bumblebees, specifically that different bumblebee species

> According to data submitted by BeeWatch users, Early bumblebee is often seen on cranesbill (144 observations), cotoneaster (91), lavender (84), chive (64) and common comfrey (53). It is important to provide flowering plants throughout the season. In the next month the plants used by Early bumblebees that are likely to be flowering are: cranesbill, chive and thyme. For more information and planting advice, please go to our new and interactive page Planting for Pollinators.

Figure 4: Example of generated planting advice emailed to BeeWatch user who submitted a photo of an Early bumblebee in June

favour different plants, and that since plants flower at different times in the year consideration should be given to availability of favoured flowers throughout the season.

In future work, we will seek to evaluate the impact of the planting recommendations provided through this process. One possibility is to analyze whether planting recommendations lead to better motivation among citizen scientists, perhaps by measuring any increase in participation in BeeWatch. Another possibility is to follow through with BeeWatch users to survey whether they have changed their gardening habits as a result of these recommendations.

ACKNOWLEDGEMENTS

Planting for Pollinators was supported by the Natural Environment Research Council (grant number: Public Engagement Pilot 2016_094) and BeeWatch was supported by Research Council UK (grant reference EP/G066051/1).

REFERENCES

[1] Michael W Berry, Murray Browne, Amy N Langville, V Paul Pauca, and Robert J Plemmons. 2007. Algorithms and applications for approximate nonnegative matrix factorization. *Computational statistics & data analysis* 52, 1 (2007), 155–173.
[2] Steven Blake, Advaith Siddharthan, Hien Nguyen, Nirwan Sharma, Anne-Marie Robinson, Elaine O'Mahony, Ben Darvill, Christopher Stuart Mellish, and Rene Van Der Wal. 2012. Natural language generation for nature conservation: Automating feedback to help volunteers identify bumblebee species. *COLING 2012:* (2012).
[3] J. Bobadilla, F. Ortega, A. Hernando, and A. GutiÃľrrez. 2013. Recommender systems survey. *Knowledge-Based Systems* 46 (7 2013), 109–132.
[4] Moody Chu, Fasma Diele, Robert Plemmons, and Stefania Ragni. 2004. Optimality, computation, and interpretation of nonnegative matrix factorizations. In *SIAM Journal on Matrix Analysis*. Citeseer.
[5] Christian Desrosiers and George Karypis. 2011. A comprehensive survey of neighborhood-based recommendation methods. In *Recommender systems handbook*. Springer, 107–144.
[6] Justin Dillon. 2016. 50 Years of JBE: From Science and Environmental Education to Civic Science. *Journal of Biological Education* 50, 2 (2016), 120–122.
[7] Michael D Ekstrand, John T Riedl, and Joseph A Konstan. 2011. Collaborative filtering recommender systems. *Foundations and Trends in Human-Computer Interaction* 4, 2 (2011), 81–173.
[8] C. Gena, R. Brogi, F. Cena, and F. Vernero. 2011. *The impact of rating scales on user's rating behavior.* Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol. 6787 LNCS. 123–134 pages.
[9] Yehuda Koren. 2009. The Bellkor Solution to the Netflix Grand Prize. *Netflix prize documentation* 81 (2009).
[10] Yehuda Koren, Robert Bell, Chris Volinsky, et al. 2009. Matrix factorization techniques for recommender systems. *Computer* 42, 8 (2009), 30–37.

[11] D. D. Lee and H. S. Seung. 1999. Learning the parts of objects by non-negative matrix factorization. *Nature* 401, 6755 (1999), 788–791.
[12] D. D. Lee and H. S. Seung. 2001. Algorithms for non-negative matrix factorization.
[13] Greg Linden, Brent Smith, and Jeremy York. 2003. Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet computing* 7, 1 (2003), 76–80.
[14] Pasquale Lops, Marco Gemmis, and Giovanni Semeraro. 2011. Recommender Systems Handbook. Content-based Recommender Systems: State of the Art and Trends (2011), 73–105. http://dx.doi.org/10.1007/978-0-387-85820-3_3
[15] Juliet L Osborne, Andrew P Martin, Chris R Shortall, Alan D Todd, Dave Goulson, Mairi E Knight, Roddy J Hale, and Roy A Sanderson. 2008. Quantifying and comparing bumblebee nest densities in gardens and countryside habitats. *Journal of Applied Ecology* 45, 3 (2008), 784–792.
[16] Jennifer Shirk, Heidi Ballard, Candie Wilderman, Tina Phillips, Andrea Wiggins, Rebecca Jordan, Ellen McCallie, Matthew Minarchek, Bruce Lewenstein, Marianne Krasny, et al. 2012. Public participation in scientific research: a framework for deliberate design. *Ecology and Society* 17, 2 (2012).
[17] Advaith Siddharthan, Christopher Lambin, Anne-Marie Robinson, Nirwan Sharma, Richard Comont, Elaine O'mahony, Chris Mellish, and René Van Der Wal. 2016. Crowdsourcing without a crowd: Reliable online species identification using Bayesian models to minimize crowd size. *ACM Transactions on Intelligent Systems and Technology (TIST)* 7, 4 (2016), 45.
[18] Jonathan Silvertown, Martin Harvey, Richard Greenwood, Mike Dodd, Jon Rosewell, Tony Rebelo, Janice Ansine, and Kevin McConway. 2015. Crowdsourcing the identification of organisms: A case-study of iSpot. *ZooKeys* 480 (2015), 125.
[19] Xiaoyuan Su and Taghi M Khoshgoftaar. 2009. A survey of collaborative filtering techniques. *Advances in artificial intelligence* 2009 (2009), 4.
[20] René Van der Wal, Helen Anderson, Annie Robinson, Nirwan Sharma, Chris Mellish, Stuart Roberts, Ben Darvill, and Advaith Siddharthan. 2015. Mapping species distributions: A comparison of skilled naturalist and lay citizen science recording. *Ambio* 44, 4 (2015), 584–600.
[21] René Wal, Nirwan Sharma, Chris Mellish, Annie Robinson, and Advaith Siddharthan. 2016. The role of automated feedback in training and retaining biological recorders for citizen science. *Conservation Biology* 30, 3 (2016), 550–561.
[22] Arjen EJ Wals, Michael Brody, Justin Dillon, and Robert B Stevenson. 2014. Convergence between science and environmental education. *Science* 344, 6184 (2014), 583–584.
[23] Ke Zhou, Shuang-Hong Yang, and Hongyuan Zha. 2011. Functional matrix factorizations for cold-start recommendation. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*. ACM, 315–324.