Sustainability at Scale: Towards Bridging the Intention-Behavior Gap with Sustainable Recommendations

Sabina Tomkins
UC Santa Cruz
satomkin@ucsc.edu

Steven Isley
Amazon
isley@amazon.com

Ben London
Amazon
blondon@amazon.com

Lise Getoor
UC Santa Cruz
getoor@soe.ucsc.edu

ABSTRACT

Finding sustainable products and evaluating their claims is a significant barrier facing sustainability-minded customers. Tools that reduce both these burdens are likely to boost the sale of sustainable products. However, it is difficult to determine the sustainability characteristics of these products — there are a variety of certifications and definitions of sustainability, and quality labeling requires input from domain experts. In this paper, we propose a flexible probabilistic framework that uses domain knowledge to identify sustainable products and customers, and uses these labels to predict customer purchases. We evaluate our approach on grocery items from the Amazon catalog. Our proposed approach outperforms established recommender system models in predicting future purchases while jointly inferring sustainability scores for customers and products.

CCS CONCEPTS

• **Information systems → Recommender systems**; • **Social and professional topics → Sustainability**; • **Computing methodologies →** *Statistical relational learning*; *Latent variable models*;

KEYWORDS

Sustainability; Probabilistic programming

1 INTRODUCTION

Living *sustainably* has become a lifestyle goal for many. However, interest in sustainability does not necessarily translate into sustainable purchases [9]. The difference between one's intentions and ability to act in line with them is referred to as the *intention-behavior*, *attitude-behavior*, or *value-action gap* [14].

Barriers to sustainable shopping include: cost, availability, skepticism of labels and insufficient marketing [9]. Knowledge about what makes a product sustainable can be a barrier to consumers [13], while Tanner and Wölfing-Kast [17] found that possessing actionable knowledge is correlated with taking action. Vermeir and Verbeke [20] found positive desires to purchase sustainable food items were impeded by a lack of certainty about those items' sustainable characteristics.

Given knowledge of sustainable products, a recommender system might offer products with these characteristics to interested customers. Ekstrand and Willemsen [4] argue that recommender systems have the potential to make recommendations that lead users towards the behaviors they wish to achieve, rather than simply reinforcing observed behavior. In this line, Starke et al. [16] show that recommender systems can be useful in producing positive energy-savings behavior. Our work is a first step towards a system to bridge the intention-behavior gap by demonstrating a method to provide consumers with accurate product-level sustainability information, a necessary component of any such system.

Defining what makes a product sustainable is not straightforward. For those who value sustainability, products can be appraised according to complex factors including: environmental impact, impact on the local economy, animal welfare considerations and benefit to the consumer [12]. For this research, we have purposefully used a broad definition of sustainability that covers all these aspects. A product which scores strongly according to any one dimension might be labeled as sustainable, and multiple sustainability-related features will increase its likelihood of being labeled as sustainable. Future research could use similar methods with more narrow definitions of sustainability.

We propose a probabilistic approach which fuses multiple weak signals to infer the sustainability of products. Here we investigate three types of signals: freely available domain knowledge, product metadata and the purchasing patterns of customers predicted to be sustainability-minded. This approach offers several advantages: we are able to flexibly incorporate prior knowledge about what might imply the sustainability of products or customers; no sustainability ground truth labels are required; and multiple recommender system inputs can be used to improve predictions. Additionally, we are able to benefit from the joint formulation of all three tasks of: discovering product sustainability scores, discovering the sustainability-mindedness of customers and inferring future purchases. We demonstrate these benefits by improving predictions of future purchases by 80.8% in precision@5 over a SVD++ [10].

RecSys '18, October 2–7, 2018, Vancouver, BC, Canada

ACM ISBN 978-1-4503-5901-6/18/10…$15.00
https://doi.org/10.1145/3240323.3240411

2 PROBLEM DEFINITION

We are given customer-item purchase data $X = (x_{i,j,d})$, where $x_{i,j,d}$ is 1 if customer c_i purchased item p_j on date d, and 0 otherwise. Our goal is to infer future purchases Y. This is an implicit feedback setting where the only information we have is which items a customer bought and when.

Our goal is to predict which items are sustainable and which customers are sustainably-minded. To that end, we will predict scores s^c and s^p, respectively, to each customer and item, corresponding to their degree of sustainability. For a customer, this score represents the extent to which they might be interested in sustainable products. For a product, the score is the extent to which this product might be environmentally sustainable. Unlike the purchases, there are no true labels for s^c or s^p, the extent to which a customer is interested in sustainability is ultimately an unknown quantity. Thus we characterize these variables with a latent formulation and infer their values without ground-truth scores.

3 OUR APPROACH

We propose a collective probabilistic model which allows us to both predict future purchases $\boldsymbol{y}$ and jointly discover sustainability scores, s^c and s^p. We construct this model with Probabilistic Soft Logic (PSL) [1], a probabilistic programming framework which offers several advantages for this setting: logical rules can intuitively capture domain knowledge, collective inference of both purchases and scores is highly efficient and latent variables facilitate the fusing of multiple signals. Next, we introduce PSL in more detail.

3.1 Probabilistic Soft Logic

In PSL relationships between variables are encoded with weighted logical rules. These rules can capture dependencies not only from observed features to target variables, but *between* target variables. This expressivity allows us to encode both domain knowledge about product sustainability, as well as product and customer similarities. Finally, PSL provides an intuitive framework for representing latent abstractions and an efficient procedure for inferring their values.

To demonstrate how PSL can be used in a recommender system setting, consider a rule which states that if two customers are similar, and one customer is sustainable, the other one is as well. We introduce a predicate Similar, which takes two customer IDs as arguments and which expresses the similarity between these two customers as a value between 0 and 1. There are many ways to express similarity, and it is possible to use multiple definitions in a single PSL model. Here we calculate similarity from latent factors learned from a SVD++ model. To express the sustainability proclivity of a customer, we introduce the predicate SustainableCustomer, which takes a customer ID, c_i, as an argument, and whose truth value corresponds to the sustainability score s^{c_i}. Together a predicate and its arguments form a logical atom, but, unlike in Boolean logic, PSL atoms can assume soft truth values in $[0, 1]$. With these predicates, and a weight w_{sim} which reflects the relative importance of this rule, we define our rule in PSL as follows:

$$w_{sim} : \textsc{Similar}(c_i, c_j) \wedge \textsc{SustainableCustomer}(c_i) \Rightarrow \textsc{SustainableCustomer}(c_j).$$

Combined with data, a PSL model defines a joint probability distribution over scores and purchases. This distribution is expressed with a *hinge-loss* Markov random field (HL-MRF) [1], a general class of conditional, continuous probabilistic graphical models. HL-MRFs provide the advantage of both high efficiency and expressivity. To perform weight-learning in the presence of latent variables, we use the method described by Bach et al. [2]. PSL is adept at fusing multiple sources of information. Next, we introduce our three sources of sustainability signals.

3.2 Three Signals of Sustainability

We propose three sources of information about what makes a product sustainable. One source is freely available information which can be collected from published material on or offline, as well as from domain experts. Another source is the product metadata. Finally, customers' purchasing patterns can be leveraged to identify additional products.

Domain Knowledge With the abundance of online publishing dedicated to sustainability, from magazines such as Mother Earth to consumer services such as GoodGuide, there are multiple sources for identifying potentially sustainable products. We incorporate knowledge of which brands are sustainable into our model.

Metadata One source of information is product metadata. This category can include a range of features, from product descriptions to the number of organic ingredients in a prepared food item. We consider two sources: certifications and specialities. Certifications are third party assessment of some sustainability-related attributes e.g., the USDA Organic program certifies that agricultural products have been produced using approved methods. Specialties are tags that provide product filters, for example: Organic, or Gluten-Free.

Sustainability-Minded Shoppers If we can identify a group of sustainability-minded shoppers then we can learn from their purchasing patterns. These customers may purchase products which are not officially certified, but which they believe to be sustainable, and which we would like to score as such. Thus, we use purchasing patterns to identify sustainable customers and additional sustainable products from their shopping.

Each type of information suffers from its own drawbacks. Domain knowledge will inevitably be sparse, covering only small subsets of large datasets. Product metadata can contain errors or omissions that make it unreliable at times. Finally, purchasing patterns are also weak signals. From an analysis point of view, it would be ideal for sustainability-minded shoppers to only purchase sustainably. However, this is unrealistic [6]. By fusing these signals, we overcome the drawbacks of each and gain a more complete understanding of what makes a product sustainable.

3.3 PSL Sustainable Discovery Model

We infer the values of three latent variables: SustainableCustomer(C), SustainableItem(P) and PredictPurchase(C, P, D). We also infer the value of the target variable Purchase(C, P, D). The values of SustainableCustomer(C), and SustainableItem(P), correspond to s^c and s^p, and reflect the sustainability scores of customers and items. To predict if a customer will purchase a product on a given day, we use PredictPurchase. We consider each purchase event as a separate random variable, Purchase(C, P, D) which takes a value in [0,1], and is 1 if customer C purchased product P on date D.

w_{np}	$:\neg\textsc{Purchase}(C, P, D)$
w_{nsc}	$:\neg\textsc{SustainableCustomer}(C)$
w_{nsp}	$:\neg\textsc{SustainableProduct}(P)$

Table 1: These rules represent prior beliefs.

In large catalogs it is common that most products won't be purchased. We reflect this with the first rule in Table 1. We also encode that we expect most customers and products to not be sustainable in the next two lines. These rules encode our prior beliefs. However, the weights to these initial priors can be updated with data.

w_i	$: \textsc{Certification}(P, Cert_i) \Rightarrow \textsc{SustainableProduct}(P)$
w_j	$: \textsc{Specialty}(P, Spec_j) \Rightarrow \textsc{SustainableProduct}(P)$
w_b	$: \textsc{SustainableBrand}(B) \wedge \textsc{Brand}(P, B) \Rightarrow \textsc{SustainableProduct}(P)$
w_s	$: \textsc{SustainableProduct}(P_1) \wedge \textsc{SimilarProducts}(P_1, P_2) \Rightarrow \textsc{SustainableProduct}(P_2)$

Table 2: *Sustainable Products*

The first rule in Table 2 relates certifications to products. A certification is awarded by an external service, for example the USDA can certify a product organic. For each certification i, we instantiate a new rule with weight w_i, to learn the relative importance of certifications. Similarly, each product might be described with a Specialty, and for each specialty j, we instantiate a unique rule and learn a weight w_j. Additionally, we can apply domain knowledge about the sustainability of companies, to products with the third rule in Table 2, which states that products offered by sustainable brands are themselves sustainable. Finally, we propagate information about similar products with the final rule that says if two products are similar, and one is sustainable, the other one will be as well.

w_{sc}	$: \textsc{SustainableCustomer}(C_1) \wedge \textsc{SimilarCustomers}(C_1, C_2) \Rightarrow \textsc{SustainableCustomer}(C_2)$
w_{shc}	$: \textsc{SustainableProduct}(P) \wedge \textsc{HasPurchased}(C, P) \Rightarrow \textsc{SustainableCustomer}(C)$
w_{spc}	$: \textsc{SustainableProduct}(P) \wedge \textsc{PredictPurchase}(C, P, D) \Rightarrow \textsc{SustainableCustomer}(C)$

Table 3: *Sustainable Customers*

w_{pp}	$: \textsc{HasPurchased}(C, P) \Rightarrow \textsc{PredictPurchase}(C, P, D)$
w_{tda}	$: \textsc{PredictPurchase}(C_1, P, D) \wedge \textsc{SimilarCustomers}(C_1, C_2) \Rightarrow \textsc{PredictPurchase}(C_2, P, D)$
w_{spp}	$: \textsc{PredictPurchase}(C, P_1, D) \wedge \textsc{SimilarItems}(P_1, P_2) \Rightarrow \textsc{PredictPurchase}(C, P_2, D)$
w_{svd}	$: \textsc{SVDPredicts}(C, P) \Rightarrow \textsc{PredictPurchase}(C, P, D)$
w_{ppp}	$: \textsc{PredictPurchase}(C, P, D) \Rightarrow \textsc{Purchase}(C, P, D)$

Table 4: *Predicting Purchases*

Next, in Table 3, we identify sustainability-minded shoppers. We first leverage similarities across customers; if two customers are similar and one is sustainable, the other one likely is as well. Additionally, if at any time a customer has purchased a sustainable product, that customer might be sustainability-minded. Finally, we model that if a customer will purchase a sustainable product, they are likely sustainablility-minded.

Next, we fuse existing knowledge into predictions of what a customer will purchase. In the first rule in Table 4, we predict that a purchased item will be purchased again. This rule is very context specific, but as our analysis focuses on food, a category where repeat purchases are common, this rule is appropriate. With the next two rules we again leverage customer and product similarities. We also incorporate predictions from other recommender systems. Here we utilize predictions made by a Singular Value Decomposition (SVD) algorithm, with the predicate SVDPredicts, which assumes a value between 0 and 1 according to the SVD algorithm. We then apply the predictions to the target variable Purchase(C, P, D), with the last rule in Table 4.

4 QUANTITATIVE EVALUATION

In the quantitative evaluation, we assess our framework by its ability to correctly predict customer purchases. To do so we compare our approach to one baseline and one state-of-the-art approach. Our approach is also notable in its ability to discover sustainable products and customers, as we have no ground truth labels for this task, we present a qualitative evaluation in Section 5.

4.1 Data

In these experiments we consider customer purchase data from Amazon.com. We focus the experiments on the Grocery category. This excludes related products such as Amazon Pantry. We choose food as a first exploration of sustainability, as there are clear sustainable metadata, such as organic and fair-trade. We create training and test sets with 10,000 customers in training, 5,000 customers in the validation set and 5,000 customers in the test set. For each customer, all purchases in the validation set occur *after* all purchases in the training set, and all purchases in the test set occur after all purchases in the validation set. Since we have implicit preference feedback (i.e., we only know what was purchased, not what was considered but not purchased), we sample 100 negative purchases for training as in Said and Bellogín [15]. To further refine the problem, we only considered products purchased below a given threshold across a large number of customers. The total number of products was approximately 21,000. In the validation and test set, for each customer, we infer purchases on a single date D.

Our external information about which companies are sustainable was collected from two online sources.[1] When filtering for the companies present in our dataset we were left with sixteen sustainable companies. These were largely food companies, with the exception of Seventh Generation which sells household products.

The specialities and certifications are provided by Amazon. We consider the following specialties: organic, organic & whole grain, all natural, gluten free, wheat free, dairy free, natural ingredients only, sustainably caught, biodegradable and not-tested-on-animals. We expect these to indicate sustainability to varying degrees, for example, *organic* should be a stronger indicator than *all natural*. The certifications included in this analysis were: non-GMO, Rainforest Alliance, all organic certifications and all fair trade certifications.

4.2 Experiments

We compare our approach to two common baselines: a nearest neighbor (NN) search and SVD++ [10] which is preferable in the implicit feedback setting. Both methods are implemented in the python package Surprise [8]. In Table 5 we show the percent improvement of the NN and our approach (PSL) over the SVD++ method, on the task of predicting purchases.

Here we see that PSL achieves the best performance according to the precision@k and Mean Average Precision (MAP). The SVD++ outperforms the NN baseline only according to the MAP. Next

[1] https://eating-made-easy.com/sustainable-food-companies/, https://www.motherearthliving.com/food-for-health/sustainable-food-companies-zmoz12jazmel

	PSL	NN
p@5	**80.8**	55.5
p@20	**48.6**	30.5
p@50	**14.9**	7.46
MAP	**63.7**	-5.80

Table 5: Percent relative improvement w.r.t. SVD++, with largest statistically significant improvements bolded.

we inspect the sustainable signals: domain knowledge, product metadata and purchasing patterns.

4.3 Effects of each signal

Here, we inspect how many products were discovered with each source of information: domain knowledge, product metadata and purchasing patterns. As our model is collective, and infers sustainability jointly, to resolve this question we ask how many products were discovered without each type of information. That is, how many products without metadata were determined to be sustainable, how many products without domain knowledge, and how many products without either were determined to be sustainable.

	Total	Domain, No Metadata	Metadata, No Domain	With Neither
Found	$\frac{862}{21000} \approx 4.1\%$	$\frac{39}{862} \approx 4.5\%$	$\frac{26}{862} \approx 3.0\%$	$\frac{13}{862} \approx 1.5\%$

Table 6: Most products were discovered using all three signals.

In Table 6, we see that 862, or roughly 4% of all products considered, were found to be sustainable. Comparing the importance of domain knowledge and metadata, we see that they are relatively similar. Of the discovered sustainable items, roughly 4.5% and 3.0% were missing metadata and domain knowledge respectively. This suggests that domain knowledge may be a stronger signal than metadata. However, only 1.5% were missing both. This shows that the combined sources of information are complementing each other rather than finding distinct sets of products. However, there is still a question of the quality of these discovered products. Next, we turn to a human evaluation to assess the discovered products.

5 QUALITATIVE EVALUATION

To assess the sustainability of discovered products, we ask a domain expert to evaluate a subset of selected products with high sustainability scores. In this evaluation, each item is scored as: Reasonably Sustainable, Partially Sustainable, or Not Sustainable. In this assessment we do not ask if it would be more sustainable to not purchase this item. For example, we assess the sustainability of a particular product of bottled water and if the bottle is made of recycled material it may obtain a high score. However, it would be more sustainable to buy a reusable water bottle. Thus, the evaluations of the products are made in isolation, without considering if a better alternative would be to not purchase at all.

We split the evaluation so that some products from each category in Table 6 are explored. Twenty products missing no information, ten products with no metadata, ten products with no domain knowledge, and ten products with neither domain knowledge or metadata are evaluated. In total 50 products are evaluated in Fig. 1.

We see that the majority of evaluated products are found to be reasonably sustainable. Both metadata and domain knowledge are useful in detecting sustainable products. However, these two categories do not cover the entire dataset. Of the products without

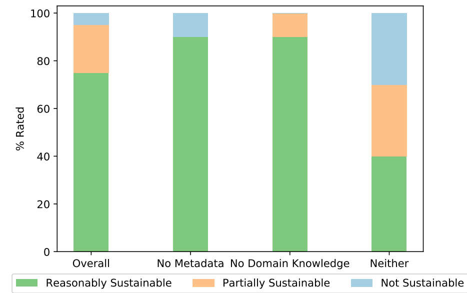

Figure 1: Human ratings for a subset of discovered products.

either metadata or domain knowledge, 40% of those predicted to be sustainable were deemed reasonably sustainable by an expert and 30% were deemed partially sustainable. While improvements can be made to better discover products with incomplete information, the majority of these products were somewhat sustainable.

6 DISCUSSION

We propose this work as a first step towards sustainable recommender systems. By discovering products of interest to sustainability-minded customers we can start to address the lack of knowledge which frustrates the behavior-intention gap. Our proposed model is able to find a reasonable number of sustainable products and the majority of these products are deemed reasonably sustainable by a domain expert. There are many ways in which this work could be expanded. In the current implementation, a small number of companies proved to be helpful in discovering sustainable products. Thus, expanding this number might improve results. Furthermore, our model would surely benefit from incorporating additional predictions from other recommender systems, especially state-of-the-art systems, which has proved useful in related work [11]. Another next step would be to utilize temporal information such as seasonal purchasing habits.

Our approach could be considered a hybrid recommender system [3, 7]. Thus, our work is similar to approaches which incorporate item features and domain knowledge with collaborative filtering [18, 19]. One consideration in our setting is that sustainability-minded customers may have constraints about what they will not buy; for example, products containing toxic pollutants. Like constraint-based methods [5], our probabilistic method can also incorporate hard constraints.

7 CONCLUSION

The intention-behavior gap expresses that people do not always behave in accordance to their intentions. This gap is partly due to the lack of easily accessible and credible information on product sustainability. To address this issue, we propose a model which utilizes three sources of information to discover sustainable products and make sustainability informed recommendations. We demonstrate that this approach leads to better recommendations than baselines. Furthermore, 74% of the discovered products are deemed reasonably sustainable by an expert human evaluator.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant Numbers CCF-1740850 and IIS-1703331.

REFERENCES

[1] Stephen H Bach, Matthias Broecheler, Bert Huang, and Lise Getoor. 2017. Hinge-Loss Markov Random Fields and Probabilistic Soft Logic. *Journal of Machine Learning Research (JMLR)* (2017).

[2] Stephen H. Bach, Bert Huang, Jordan Boyd-Graber, and Lise Getoor. 2015. Paired-Dual Learning for Fast Training of Latent Variable Hinge-Loss MRFs. In *International Conference on Machine Learning (ICML)*.

[3] Robin Burke. 2007. *Hybrid Web Recommender Systems*.

[4] Michael D. Ekstrand and Martijn C. Willemsen. 2016. Behaviorism is Not Enough: Better Recommendations Through Listening to Users. In *Conference on Recommender Systems(RecSys)*.

[5] A. Felfernig and R. Burke. 2008. Constraint-based Recommender Systems: Technologies and Research Issues. In *Proceedings of the 10th International Conference on Electronic Commerce (ICEC '08)*.

[6] Andrew Gilg, Stewart Barr, and Nicholas Ford. 2005. Green consumption or sustainable lifestyles? Identifying the sustainable consumer. *Futures* 37, 6 (2005).

[7] F. S. Gohari and M.J. Tarokh. 2017. Classification and Comparison of the Hybrid Collaborative Filtering Systems. *International Journal of Research in Industrial Engineering* 6, 2 (2017).

[8] Nicolas Hug. 2017. Surprise, a Python library for recommender systems. http://surpriselib.com. (2017).

[9] Renée Shaw Hughner, Pierre McDonagh, Andrea Prothero, Clifford J. Shultz, and Julie Stanton. 2007. Who are organic food consumers? A compilation and review of why people purchase organic food. *Journal of Consumer Behaviour* 6, 2-3 (2007).

[10] Yehuda Koren. 2008. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Conference on Knowledge discovery and data mining (SIGKDD)*.

[11] Pigi Kouki, Shobeir Fakhraei, James Foulds, Magdalini Eirinaki, and Lise Getoor. 2015. Hyper: A flexible and extensible probabilistic framework for hybrid recommender systems. In *Conference on Recommender Systems (RecSys)*.

[12] Yuko Onozaka, Gretchen Nurse, and Dawn Thilmany McFadden. 2011. Defining Sustainable Food Market Segments: Do Motivations and Values Vary by Shopping Locale? *American Journal of Agricultural Economics* 93, 2 (2011).

[13] Josephine Pickett-Baker and Ritsuko Ozaki. 2008. Pro-environmental products: marketing influence on consumer purchase decision. *Journal of Consumer Marketing* 25, 5 (2008).

[14] Sustainable Consumption Roundtable. 2006. I will if you will: Towards sustainable consumption. (2006).

[15] Alan Said and Alejandro Bellogín. 2014. Comparative Recommender System Evaluation: Benchmarking Recommendation Frameworks. In *Conference on Recommender Systems (RecSys)*.

[16] Alain Starke, Martijn Willemsen, and Chris Snijders. 2017. Effective User Interface Designs to Increase Energy-efficient Behavior in a Rasch-based Energy Recommender System. In *Conference on Recommender Systems (RecSys)*.

[17] Carmen Tanner and Sybille Wölfing Kast. 2003. Promoting sustainable consumption: Determinants of green purchases by Swiss consumers. *Psychology & Marketing* 20, 10 (2003).

[18] B. Towle and C. Quinn. 2000. Knowledge Based Recommender Systems Using Explicit User Models. In *AAAI Technical Report*.

[19] Abdulbaki Uzun, Christian Räck, and Fabian Steinert. 2010. Targeting More Relevant, Contextual Recommendations by Exploiting Domain Knowledge. In *Workshop on Information Heterogeneity and Fusion in Recommender Systems (HetRec)*.

[20] Iris Vermeir and Wim Verbeke. 2006. Sustainable food consumption: Exploring the consumer attitude–behavioral intention gap. *Journal of Agricultural and Environmental ethics* 19, 2 (2006).