Blending Machine and Human Learning Processes

Kevin Crowston
Syracuse University
crowston@syr.edu

Carsten Østerlund
Syracuse University
costerlu@syr.edu

Tae Kyoung Lee
University of Utah
tl452@cornell.edu

Abstract

Citizen science projects rely on contributions from volunteers to achieve their scientific goals and so face a dilemma: providing volunteers with explicit training might increase the quality of contributions, but at the cost of losing the work done by newcomers during the training period, which for many is the only work they will contribute to the project. Based on research in cognitive science on how humans learn to classify images, we have designed an approach to use machine learning to guide the presentation of tasks to newcomers that help them more quickly learn how to do the image classification task while still contributing to the work of the project. A Bayesian model for tracking this learning is presented.

1 Introduction

Online production communities rely on participants to contribute and in many cases to manage and maintain the community. To be successful, communities need to sustain a critical mass of skilled and active participants [9, 16], which requires attracting newcomers and helping them learn to become effective participants in the community.

In traditional organizations, new members often go through formal training to learn how to contribute. However, online communities present a challenge to newcomer orientation and training. Many online groups consist of volunteers contributing in their free time, reducing their willingness to participate in formal training regimes prior to engaging with the community. A further complication is the skewed distribution of contributions in many projects: most volunteers contribute only a few times and only a few become long-term contributors. As a result, increasing the barrier to entry and delaying newcomers' contributions (e.g., by requiring training) may mean that many end up not contributing at all.

Some crowdsourcing systems allow newcomers to learn through observation of the contributions of more experienced users. For instance, Bryant et al. [3] found in a study of Wikipedia that new editors begin by reading articles before they make their initial contribution. However, this form of transparency is not possible for all types of online work and it can also take significant time for newcomers to learn through observation.

To make online communities more effective calls for new approaches to newcomer learning that redefine the relationship between the humans and the infrastructure. The technology must enable motivated participants to make productive contributions to the community while also supporting an efficient and engaging learning process for newcomers.

In this paper, we present the design of a citizen science project site that incorporates machine learning to guide training for new volunteers. Citizen science is a broad term describing scientific projects that rely on contributions to advance scientific research from members of the general public (i.e., citizens in the broadest sense of the term). There are several different kinds of citizen science projects: some have volunteers collect data, while others, including the ones we examine in this paper, have volunteers analyze already collected data. The interactions between volunteers and the project organizers are typically via the Web, e.g. on a site that accepts contributed data or that presents data to be analyzed and collects volunteers' data or annotations (e.g., Zooniverse.org).

Many online citizen science projects just give volunteers a brief overview of the task and the site features before allowing them to contribute. This approach has some advantages. First, it ensures that more of the volunteers' time is being used for the work of the project. Furthermore, knowing that the work is useful and being given challenging tasks may be motivating for volunteers. However, if it takes time to learn to do the task correctly, then the initial contributions may not be of high enough quality to be useful for science (as experienced by [4]). Furthermore, if new volunteers find the task too challenging, they may become discouraged and drop out of the project.

To ensure that volunteers understand the task, a few projects (e.g., Star Dust @ Home) provide explicit training for new users, as would a traditional organization. A disadvantage of this approach is that during the training newcomers are not being productive and many might never do any real work. Furthermore, developing a training program requires additional work by the project developers to create appropriate training materials.

In short, citizen science projects face a dilemma in how to handle newcomers. Providing training might increase the quality of contributions, but at the cost of the work done by newcomers during the training period, which for many is the only work they will contribute. On the other hand, not providing training might mean that the initial contributions are not useful. Our system addresses this dilemma.

2 Theory

The design of our system draws on cognitive theories about how humans learn to classify, leading to insights about how a system can train users and track human performance to estimate a person's ability at the task. We focus in particular on theories about image classification, which is a common citizen science data analysis task, and the specific focus of the system we are building. For example, in the Zooniverse Snapshot Serengeti project, volunteers identify the species of animals in photographs.

Cognitive theories suggest that people learn to classify images though exposure to prototypes and exemplars of known categories. Prototypes serve as a heuristic: an average representation of an entire category [12]. Exemplars function as examples for the category [13]. When individuals classify stimuli, they find similarity of stimuli with the prototypes and exemplars. Here, similarity is based on their own internal representation (i.e., psychological representation), rather than external properties of stimuli [20]. When individuals are asked to generalize a category, they evaluate several characteristics and weight each of these characteristics [e.g., 10, 18, 21, 22]. That is, individuals make a decision if a stimulus belongs to a category depending on how much the stimulus is similar with or different from the prototypes and exemplars in certain characteristics and how the certain characteristics are important in deciding similarity (i.e., weight). As individuals experience more stimuli, they update the weights for the characteristics of stimuli.

Therefore, to support learning of image classification, volunteers should be continuously provided with good prototypes and exemplar images. For example, many Zooniverse projects provide a "field guide", with example images of the kinds of objects to be classified.

To properly target training requires some estimation of a volunteer's current level of knowledge. Currently, few citizen science projects evaluate volunteers' knowledge level. Those that do generally rely on proxies, such as the number of classifications contributed. To track volunteer performance, we are adapting the Bayesian Knowledge Tracing Model, proposed by Corbett and Anderson [8]. Bayesian methods are widely used to improve the performance of machine learning systems and human learning [11, 23]. In particular, the Bayesian Knowledge Tracing Model has been widely applied to model student learning. The model traces student's' knowledge or skill changes as they practice different skills.

Volunteer models can be used to provide individualized feedback on user's action. If the system can track what each individual learns, it can provide individualized feedback adjusting their level of knowledge or skills. Providing proper feedback is critical in learning process [7, 14, 17]. In an experiment, Corbalan, Kester, and Van Merriënboer [7] found that when feedback was provided for participants on their performance, they were more motivated than when feedback was not provided. In particular, explanatory feedback, explaining why their answer is correct or wrong, has been found to be more effective than corrective feedback, saying whether the answer is correct or wrong [6]. Tracking individuals' performance allows a system to provide explanatory feedback suited for their level.

The above discussion has focused on human learning of classification tasks, but machine learning for image classification is also an active research area that has recently seen great advances [e.g, 5]. There is evidence that humans and computers offer distinct skills in classification. For example, Beaumont et al. [2] created a hybrid model of machine learning combined with crowdsourced training data from citizen scientists for the Milky Way Project. They found that "untrained" citizens can identify patterns that machines cannot detect without training, while machine learning algorithms can use the output of citizen science projects as input training sets.

3 Machine-learning-supported training

To address the problems faced by citizen science projects, we are building a system that will enable a symbiotic relationship between citizen science volunteers and computer algorithms, each helping the other learn to classify images. Volunteers will sort through vast amounts of data to build a robust "gold standard" image dataset that will train machine-learning algorithms. As the ML algorithms learn from this classified dataset, they will be able to select images that assist humans to learn.

3.1 Data

In addition to a store of images to be classified, the system includes two data sets: the image-taxonomy and the gold-standard data sets. The first is the descriptions and examples of image classes. The second data store contains a subset of the images (referred to as "gold standard" data) that have been labelled by human experts with the correct classification, possibily including "none of the above" for images that do not fit any of the known classes.

3.2 Machine learning

Machine learning (ML) models are trained using the gold standard data (one model for each class of image). The trained ML models are applied to all unlabelled images, annotating each unlabelled images with the ML model's level of confidence that the image is a member of each class. Often, the confidence level for one of the classes will be much higher than for the others, suggesting that that image is a member of that class. But it also possible for none of the confidence levels to be high, meaning that the ML models are not able to classify the image or for more than one confidence to be at an intermediate level, meaning that the ML models are uncertain about the classification.

As noted above, ML models and human experts do not necessarily see the same things in data. The relation between the ML-determined degree of confidence and likelihood of the image being of the given class is expected to show a distribution as shown in Figure 1. Nearly all images above a certain threshold of ML confidence will be judged by the human experts to be of that class; nearly all below a certain threshold as not of that class; and in the intermediate range of confidence, a mix of in and not in the class.

3.3 Training citizen science volunteers

Using a citizen science platform such as Zooniverse, volunteers are presented with images and asked to classify them into one of the known categories, none of the above or "no image" for images that in fact do not include an object of interest.

Citizen science projects typically provide a brief introduction to the project, explaining its goals, how to interpret the images and how to use the classification interface. The research on learning reviewed above suggests that an effective way to train humans to perform image classification tasks is to provide them with exemplary images from which to learn. Accordingly, citizen science classification interfaces generally show the volunteers examples of images of all the classes as exemplars to guide the choice. When a classification is selected, a larger image and a brief description can be displayed to reinforce the exemplar.

As noted above, a main advance in our system is that we will use the machine learning results to train the human volunteers. Specifically, the system, guided by the ML results, will move new volunteers through a sequence of levels in which they are presented with different classification tasks intended to improve their ability to classify images [20]. Essentially, the system is acting like a tutoring system in picking tasks to help a beginner to learn, but selecting from the natural tasks of the citizen science project rather than from a predefined set of training materials.

Specifically, a volunteer who has just joined the project will be presented images that have been classified by the ML models as being likely to be of one of only two very distinctive classes. For each image, volunteers will be asked to annotate it as being an instance of one of the two classes or "none of the above" (i.e., with a reduced version of the interface). Because the ML has a high level of confidence in the classification of the images, it is most likely that these images are of the identified class. We further expect that these will be exemplary images that will further help the volunteer to learn how to identify that class of image.

It may also be desirable to give beginner volunteers a few images from the gold standard data set to classify, since knowing the correct classification makes it possible to give the volunteers feedback on the correctness of their classifications, which is also effective in promoting learning. Depending on the ML performance, it might be possible to use the ML classification as a basis for feedback. If there is a level of ML confidence above which essentially all images are in fact of the predicted class, then users could be given feedback based on those images as well.

The system will maintain a model of each volunteer's ability to classify images of each class and will update the models after each classification (e.g., increasing its estimate of the volunteers' ability when they agree with an assessment and decreasing it if they disagree). We propose using Corbett and Anderson's [8] Bayesian Knowledge Tracing model for the volunteer model, with modifications to account for the possibility that the ML classification might be incorrect, rather than the volunteer's classification. A limitation of this approach is that the training will only be available for logged in users whose progress can be tracked.

The model maintains an estimate of $p(L_n)$, the probability that the volunteer has learned how to classify after having classified n images. There is a separate model for each class of image. From [8], the formula to update the model's estimate of the volunteer's ability is equation 1 in Table 1 (below), where $p(T)$ is the probability of learning to classify if the volunteer does not already know how. Noted that the model does not include forgetting, though that could be added. $p(L_n)$, the updated probability that the volunteer knows how to classify given their answer (agreeing or disagreeing with the ML classification) is estimated using Bayesian inference from the prior estimate and the answer, as shown in equation 2 [1]. $p(L_0)$, the initial estimate of the volunteer's ability, is another parameter of the model.

The components of those equations are given in equations 3–5. In those equations, $p(M)$ is the estimated probability that the particular image seen on this step is of the class identified by the ML model. This factor is novel in our system. There are two further parameters drawn from [8]: $p(G)$, the probability of a volunteer getting the answer right without knowing how to classify (guessing) and $p(S)$, the probability of getting the answer wrong even while knowing how to classify (slipping). Note that a volunteer's answers being right and wrong are defined according the image's true classification.

The chance of the volunteer agreeing with the ML result while knowing how to classify is thus just the chance that the ML is correct and the volunteer has not slipped or that the ML is incorrect and the volunteer slipped (equation 3). The probability of the volunteer agreeing with the ML result unconditioned is the probability that both the ML and the human are correct or both are incorrect (equation 4). Finally, the probability that the volunteer has a correct evaluation is that they know how to classify and did not slip or do not know but guessed correctly (equation 5). The formula for the case of the volunteer disagreeing with the ML model (equation 6) is just the inverse: since agreeing and disagreeing are binary decisions, the probability of disagreeing is one minus the probability of agreeing.

The same model can be used to predict a volunteer's classification of a image given the answers on previous classifications (i.e., using equation 4). The parameters, $p(T)$, $p(G)$, $p(S)$ and initial ability, $p(L_0)$, can be estimated by fitting the model to minimize the prediction error for an initial dataset. The same parameters can be used for all classes of image, reducing the number of parameters to be estimated, or, with enough data, different parameters can be estimated for each class of image (e.g., to allow some classes to be harder to learn or easier to confuse). More advanced approaches to estimation have been suggested that take into account features of the answer in estimating the probability of a slip or guess [1] or to estimate models with parameters individualized for each student [24].

Once estimated, the model can be used to track a new volunteer's learning. When the volunteer model shows that the volunteer's abilities on the set of classes being trained are above a certain threshold, the volunteer will be advanced to the next training level, in which they see images believed by the ML to be of additional classes. Corbett and Anderson [8] used a threshold of 0.95, though without specific justification. Again, during the training period, volunteers will only see images that the ML model has classified with high confidence, which should serve as good exemplars from which to learn the additional classes.

A quick simulation of the model given above with $p(T) = 0.2$ and $p(L_0) = 0.3$ shows that if the volunteer agrees with the model on each step, they will reach the 0.95 level after only 3 steps when given images that are at least 0.95 likely to be of the given class. With images that are at least 0.8 or 0.9 likely, the process takes 4 steps. Of course, volunteers may not always agree with the ML if they are still learning to classify or slip. In [1], the baseline probability of a slip was 44% and of a guess, 6.6%. While it is unlikely that these numbers apply exactly to the citizen science tasks, using the parameters in the simulation and allowing for occasional disagreement raises the median number of classifications needed in each condition by 1, though the learning process is occasionally extended.

The progression through the levels is also expected to be motivating for volunteers, as it will appeal to their sense of accomplishment. This motivation can be further emphasized in the interface, e.g., by showing the additional classifications to be presented in the future greyed out or with a lock icon and with appropriate messaging when mastery at the current level is achieved.

Once the user has completed all of the rounds of training on the different classes of images, they are considered fully qualified and will be given images to classify at varying levels of ML certainty in all known classes or even images for which the ML has no good classification, thus further contributing to the work of the project. Since the system is tracking each volunteer's ability, it can also assign tasks based on ability (e.g., assigning harder tasks to more capable volunteers).

Table 1. Model for volunteer learning and image classification.

1)	$p(L_n) = p(L_{n-1} \vert \text{answer}) + \left(1 - p(L_{n-1} \vert \text{answer})\right) p(T)$	
2)	$p(L_{n-1} \vert \text{agree}) = \dfrac{p(\text{agree} \vert L_{n-1})\, p(L_{n-1})}{p(\text{agree})}$	
3)	$p(\text{agree} \vert L_{n-1}) = p(M_{n-1})\left(1 - p(S)\right) + \left(1 - p(M_{n-1})\right) p(S)$	
4)	$p(\text{agree}) = p(M_{n-1})\, p(\text{correct}) + \left(1 - p(M_{n-1})\right)\left(1 - p(\text{correct})\right)$	
5)	$p(\text{correct}) = p(L_{n-1})\left(1 - p(S)\right) + \left(1 - p(L_{n-1})\right) p(G)$	
6)	$p(L_{n-1} \vert \text{disagree}) = \dfrac{(1 - p(\text{agree} \vert L_{n-1}))\, p(L_{n-1})}{(1 - p(\text{agree}))}$	
7)	$p(M_n) = p(M_{n-1} \vert \text{agree}) = \dfrac{p(\text{agree} \vert M_{n-1})\, p(M_{n-1})}{p(\text{agree})}$	
8)	$p(M_{n-1} \vert \text{disagree}) = \dfrac{(1 - p(\text{agree} \vert M_{n-1}))\, p(M_{n-1})}{(1 - p(\text{agree}))}$	
9)	$p(\text{agree} \vert M_{n-1}) = p(\text{correct})$	

Model parameters

$p(L)$	volunteer knows how to classify
$p(T)$	volunteer learns how to classify on this step
$p(M)$	ML classification is correct
$p(S)$	volunteer classifies incorrectly even though they know how (slip)
$p(G)$	volunteer classifies correctly even though do not know how (guess)
n	number of classifications

3.4 Image classification

The system uses judgement from multiple volunteers to make the final decisions on classification of images. Explicitly modelling the level of confidence in the classification of an image should make much more efficient use of human effort than the usual approach of having each item looked at by as many as fifteen volunteers to find a consensus. We anticipate that images may be classified with only a few human classifications if the ML confidence is high and the volunteers agree.

To do so, the system maintains a model of the likely classification of each image that is initialized by the ML model (i.e., $p(M_0)$). The ML confidence could be used directly, or adjusted to reflect a probability based on the curves shown in Figure 1.

Each human judgement is used to update the beliefs, as shown in equations 7–9. In this case, n is also the number of classifications, but in this case, the number of classification of the image done by different volunteers. Note that this model takes in to account differences in volunteer ability when forming a belief for the classification of images.

If the level of belief in a particular classification crosses a threshold, meaning that there is a consensus among the ML models and the human classifiers on the classification, the image can be given that label. Contrariwise, if after some number of human classifications there is no consensus, then the image can be labelled as none of the above. The process depends though on the accuracy of the human labelers. If the chance that they slip is too high (for example), it is hard to learn from their answers.

Successfully classified images will be provided to the science team to use. They can also be added to the gold standard data and used to retrain the ML model for image classification, thus using human judgement to improve the machine learning model. Indeed, the system can pick images for the volunteers to classify that will be particularly informative for improving the ML models (e.g., images that have confidence levels between the cutoffs), a process called active machine learning. However, as [15] point out, when picking an item to be classified in a crowdsourcing setting, the number of existing classifications should be considered. If the item already has many classifications, another will not reduce the ML model uncertainty. Finally, the parameters for learning model can be periodically re-estimated using the additional data.

4 Discussion

In this paper, we have presented a system that uses ML classifications of images to guide training for human volunteers in a citizen science project. The goal of the training is to help volunteers more quickly learn how to classify images and thus become productive contributors to the project. We expect that this training will also motivate users to contribute more. If the system works as expected, it will be an approach that should be of interest to other citizen science projects.

An important benefit of this approach is that because the ML cannot be certain of the classification, having a volunteer—even a beginner being trained—

confirm the classification is still useful to the project. This approach contrasts with training that is either entirely preset or that relies exclusively on gold standard data. In those cases, the work done by the volunteer as part of the training is does not directly advance the project's work. As many volunteers report that they are motivated by the fact that they are contributing to science [19], keeping the work real is important.

The system described above offers an interesting platform for experimentation. Our first planned experiment is to compare the performance of volunteers who have gone through the training process described above to the performance of those who start right away with the full set of classes for classification (i.e., the typical approach for citizen science projects). We want to test if users who get the training contribute more and show better performance on the classification tasks.

Second, the training system described above has a large number of parameters (e.g., how many and which classes to introduce at each level, the ML certainty cutoffs or the right mix of images of different certainties at different points in the process). Experimentation will be useful to determine the optimal settings. For example, we can test the benefits and tradeoffs of advancing volunteers to higher levels more quickly: quicker advancement might be good for motivation but negative for performance (and vice versa).

Finally, the system will enable us to experiment with other factors that affect volunteer performance, e.g., the kinds of motivational messages provided or information on the novelty of images. A particularly interesting set of questions are around the effects of feedback that can be provided to volunteers based on the ML certainties. Again, it is possible that there are tradeoffs involved, e.g., that letting a volunteer know what the ML evaluation was might be useful feedback to improve performance but also potentially demotivating if the ML and the volunteer disagree or volunteers feel that their contributions are unnecessary given the ML.

The main contribution of the paper has been to discuss how machine learning can be used to support learning in a citizen science project and to present a Bayesian model for tracking learning progress in this setting. The proposed system embodies a redesigned relationship between the technology of the system and the human volunteers to facilitate learning by both.

5 Acknowledgements

This work has been supported by a grant from the National Science Foundation.

6 References

[1] Ryan S. J. d. Baker, Albert T. Corbett, and Vincent Aleven, "More accurate student modeling through contextual estimation of slip and guess probabilities in Bayesian knowledge tracing," in *Proceedings of Intelligent Tutoring Systems*, 2008, pp. 406–415. doi: 10.1007/978-3-540-69132-7_44

[2] Christopher N Beaumont, Alyssa A Goodman, Sarah Kendrew, Jonathan P Williams, and Robert Simpson, "The Milky Way Project: Leveraging citizen science and machine learning to detect interstellar bubbles," *The Astrophysical Journal Supplement Series,* vol. 214, p. 3, 2014

[3] Susan L. Bryant, Andrea Forte, and Amy S. Bruckman, "Becoming Wikipedian: Transformation of participation in a collaborative online encyclopedia," in *Proceedings of GROUP Conference*, Sanibel Island, FL, 2005

[4] Roberto Bugiolacchi, Steven Bamford, Paul Tar, Neil Thacker, Ian A. Crawford, Katherine H. Joy, Peter M. Grindrod, and Chris Lintott, "The Moon Zoo citizen science project: Preliminary results for the Apollo 17 landing site," *Icarus,* vol. 271, pp. 30–48, 2016. doi: 10.1016/j.icarus.2016.01.021

[5] D. Ciregan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *Proceedings of Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012, pp. 3642-3649. doi: 10.1109/CVPR.2012.6248110

[6] Ruth C Clark and Richard E Mayer, *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*: John Wiley & Sons, 2011.

[7] Gemma Corbalan, Liesbeth Kester, and Jeroen JG van Merriënboer, "Dynamic task selection: Effects of feedback and learner control on efficiency and motivation," *Learning and Instruction,* vol. 19, pp. 455–465, 2009

[8] Albert T Corbett and John R Anderson, "Knowledge tracing: Modeling the acquisition of procedural knowledge," *User modeling and user-adapted interaction,* vol. 4, pp. 253–278, 1994

[9] Nicolas Ducheneaut, "Socialization in an open source software community: A socio-technical analysis," *Computer Supported Cooperative Work,* pp. 323–368, 2005

[10] Matt Jones, W Todd Maddox, and Bradley C Love, "Stimulus generalization in category learning," in *Proceedings of Annual Meeting of the Cognitive Science Society*, 2005, pp. 1066–1071

[11] Mohammad M. Khajah, Brett D. Roads, Robert V. Lindsey, Yun-En Liu, and Michael C. Mozer, "Designing engaging games using bayesian optimization," in *Proceedings of CHI Conference on Human Factors in Computing Systems*, Santa Clara, California, USA, 2016, pp. 5571–5582. doi: 10.1145/2858036.2858253

[12] ShinWoo Kim and Gregory L Murphy, "Ideals and category typicality," *Journal of Experimental Psychology: Learning, Memory, and Cognition,* vol. 37, p. 1092, 2011

[13] Chan Kulatunga-Moruzi, Lee R Brooks, and Geoffrey R Norman, "Teaching posttraining: influencing diagnostic strategy with instructions at test," *Journal of Experimental Psychology: Applied,* vol. 17, p. 195, 2011

[14] Detlev Leutner, "Guided discovery learning with computer-based simulation games: Effects of adaptive and non-adaptive instructional support," *Learning and Instruction,* vol. 3, pp. 113–132, 1993

[15] Christopher H Lin and Daniel S Weld, "Re-active learning: Active learning with relabeling," in *Proceedings of AAAI Conference on Artificial Intelligence*, 2016

[16] Sanna Malinen, "Understanding user participation in online communities: A systematic literature review of empirical studies," *Computers in Human Behavior,* vol. 46, pp. 228–238, 2015

[17] Roxana Moreno and Alfred Valdez, "Cognitive load and learning effects of having students organize pictures and words in multimedia environments: The role of student interactivity and feedback," *Educational Technology Research and Development,* vol. 53, pp. 35–45, 2005

[18] Robert M Nosofsky, "Attention, similarity, and the identification–categorization relationship," *Journal of experimental psychology: General,* vol. 115, p. 39, 1986

[19] M. Jordan Raddick, Georgia Bracey, Pamela L. Gay, Chris J. Lintott, Phil Murray, Kevin Schawinski, Alexander S. Szalay, and Jan Vandenberg, "Galaxy Zoo: Exploring the Motivations of Citizen Science Volunteers," *Astronomy Education Review,* vol. 9, pp. 010103–18, 2010

[20] Brett D Roads and Michael C Mozer, "Improving human-machine cooperative classification via cognitive theories of similarity," *Cognitive Science: A Multidisciplinary Journal,* In press

[21] Roger N Shepard, "Toward a universal law of generalization for psychological science," *Science,* vol. 237, pp. 1317–1323, 1987

[22] Pawan Sinha and Richard Russell, "A perceptually based comparison of image similarity metrics," *Perception,* vol. 40, pp. 1269–1281, 2011

[23] J. B. Tenenbaum, "Bayesian modeling of human concept learning," in *Advances in Neural Information Processing Systems*. vol. 11, M. Kearns, S. Solla, and D. Cohn, Eds. Cambridge, MA: MIT Press, 1999, pp. 59–65.

[24] Michael V Yudelson, Kenneth R Koedinger, and Geoffrey J Gordon, "Individualized bayesian knowledge tracing models," in *Proceedings of International Conference on Artificial Intelligence in Education*, 2013, pp. 171-180