**Introduction/Overview**

* Zooniverse engages volunteers to classify large datasets – some stats on how wildly successful this is, # projects, # volunteers, # publications
  + However, as the number of projects seeking volunteer classifiers grows, projects need to become more efficient to keep pace with data productions.
  + In fact, since launching Zooniverse’s “build your own” platform, the growth of new projects has begun to outpace the growth of the volunteer community [Fig 1- dig up old or use Helen’s for growth in projects & volunteers].
* **Mobile platforms and machine learning** provide two avenues to speeding the pace of data processing and enabling citizen science projects to keep pace with the deluge of data.
  + Mobile phones provide a way to engage more people at more opportunities throughout the day.
  + Machine learning allows for automatic image identification, thus reducing the number of images that require human input.
* However, not all projects are immediately suited to these two approaches. For example, many ecology projects ask volunteers to identify which of > 40 species is visible in an image (e.g. [www.snapshotserengeti.org](http://www.snapshotserengeti.org), [www.cameracatalogue.org](http://www.cameracatalogue.org)).
  + While good interface design can make these complex tasks accessible for untrained volunteers, the complexity of the interface does not translate well to the small screen of a mobile phone.
  + Similarly, identifying which of dozens of species is present in a photograph is a largely unsolved machine learning problem.
* We have identified an alternative approach to image identification that could allow projects to better leverage growing mobile platforms and existing machine learning techniques. The “collaborative filtering” approach converts a complex question into a series of simple questions which are presented separately to users.
  + For example, a species identification task would be broken down into a series of separate workflows consisting of a single yes/no question. The first question might be “Is there an animal in the photo?” Images that are retired as containing animals would be passed on to a second question, such as “Is the animal a whale?” Images retired as containing a whale could be passed to increasingly refined questions such as “Is the whale a humpback whale?” or “Is the whale a blue whale?” or “Is the whale doing a dance?” Each question successively “filters out” images of interest, whittling down the dataset to the images the research team most needs to extract detailed information for.
  + Simplifying a complex task into single yes/no questions not only provides more portability to mobile devices, but also enables the integration of machine learning for automated image identification at every step of the filtering process. The questions that a machine classifier is asked to answer at any stage means that the machine learning routines could be much simpler.
  + Thus, although images would ultimately require more classifications to produce the same amount of final information, the ease of classification and ability to integrate existing machine learning routines could provide benefits to efficiency that far outweigh the increase in classification counts.
* However, changing the structure and nature of a project so dramatically could have unanticipated and profound impacts on volunteer communities and classification accuracy.
  + For example, Supernova Hunters posts new data every day at 2pm and announces when the data are available. [Volunteers complete a new dataset within X minutes, meaning that only a small percentage of would-be contributors ever have the opportunity to participate, and driving the demographics to overwhelmingly older white male volunteers.]??
  + Any broader citations that might cause us to expect a change in community?
* Here we explore the potential consequences of implementing a collaborative filtering approach in terms of user contributions, engagement, and accuracy by simultaneously running both types of workflow simultaneously on an existing Zooniverse citizen science project, Snapshots at Sea.

**Methods**

Snapshots at Sea was launched June 9, 2016 to identify images containing humpback whale flukes from photographs taken by tourists and researchers onboard cruise ships. The undersides of humpback whale flukes are individually identifiable by color pattern and shape; images containing humpback whale flukes with visible undersides are passed to another citizen science project, Whales as Individuals, where volunteers help prepare the photos for individual recognition algorithms.

*Data Collection*

* We acquired 13,264 images taken by cruise passengers (?) [Need info from Ted]
* We randomly assigned these into two subsets of 6,632 images each, which were assigned to a “Survey” workflow (standard for many ecology projects) and a “Yes/No” workflow (the collaborative filtering approach, which Snapshots at Sea was originally launched as), Fig 2 (Flowchart, Helen).
* Because yes/no questions provide less information than the survey task, the collaborative filtering approach requires more classifications to complete the same number of images (see Fig 2).

*Survey*

* In the survey workflow, volunteers are presented with a list of five options: “Whale or Dolphin,” “Seal, sea otter, etc.,” “Bird,” “Fish,” and “No animal.” When they select one of these options, a pop-up screen appears with reference photos and additional details about their selection, and they can choose to select either “cancel” or “identify.”
* For “Whale or Dolphin,” users are asked two additional Yes/No questions within the pop-up screen: “Is the underside of the tail visible?” and “If YES, does the tail belong to a humpback whale?”
* Volunteers can select and identify multiple species in an image; when they are done classifying, they submit their identifications by selecting “Done” below the species list.
* Images are shown to 10 volunteers before being retired from circulation.

*Yes/No*

* In the collaborative filtering workflow, users are presented with a single yes/no question at a time. All of the 6,632 images are passed to Question 1, “Is there at least one animal in this photo? (Humans don't count!)” When all images have been seen by 5 volunteers, Question 1 is de-activated and Question 2 is activated.
* Images for which at least 4 of 5 volunteers answered “Yes” are passed to Question 2, which asks users “Are there any whales or dolphins in this photo?” Images that do not meet that threshold are retired as having no animals.
* When Question 2 is complete, images with at least 4/5 agreement of “Yes” are passed to Question 3, “Is there a fluke (tail) with its underside visible in this photo?” When Question 3 is complete, images with at least 3/5 agreement are passed to Question 4, “Is this fluke a Humpback Whale?”

We launched the experiment on June 7, and announced it via newsletter on June 8. Because Snapshots at Sea already had an active volunteer community used to the collaborative filtering approach, we sent one email to the existing community announcing new data and a new, alternative workflow, and one email to volunteers who had previously participated on another, now-retired, marine-focused project.

When Question 1 of the Yes/No workflow was completed, we activated Question 2 and sent out another newsletter announcing both new data and the alternative survey workflow. This newsletter went to all volunteers who had classified on Snapshots at Sea before or during the experiment – thus including all newly recruited volunteers.

After all four questions on the Yes/No workflow were completed, we sent out a final newsletter to the Snapshots at Sea community alerting them to the remaining data on the survey workflow.

*Data analysis*

* What I write up depends on what analyses we include…
* Retrieved all classifications from Snapshots at Sea made since the launch of the alternative workflow
* Identified users as “existing” if they had classified on the original SAS and as “new” if their first classification was during the experiment.

**Results**

Need help distilling results!

Fig 3: Instantaneous (a) and Cumulative (b) classifications through time for each approach. Cumulative classifications are plotted as proportion of total required to complete the dataset (66,320 for the survey workflow and 95,195 for the collaborative filtering workflow). Data additions are indicated in gray dashed vertical lines, newsletters in blue.

Table 1: Users, Classifications, and Talk comments (this needs updating when survey finishes)

*Um, what are the key things we see? This is a mix of results & discussion, trying to figure out key talking points*

* With this particular image set, the collaborative filtering approach required 95,195 classifications to completely classify the 6,632 images, whereas the survey approach required 66,320. (Fig 2: flowchart)
* Yes/No was more efficient. It took 20 days to completely classify 6,632 images via the Yes/No approach and 38 days to classify the same number on the survey workflow (Fig 3). While newsletters play a role in driving traffic, they don’t explain the difference in activity across the two workflows: Q3 and Q4 were completed in X days despite having no newsletters, whereas the survey workflow took an additional Y days even after an exclusively survey-focused newsletter.
* Volunteers contributed more classifications on the Yes/No questions, and, even though each individual binary classification took less time than a survey classification, volunteers actually spent more time classifying on the collaborative filtering questions.
* A surprising number of volunteers only contributed to one of the two workflow options. This held true for both volunteers who had previously classified on Snapshots at Sea as well as those who were new to the project. Volunteers who contributed to both workflows classified many more images and participated more actively on the discussion forums than those who only classified on one workflow or the other. These volunteers also classified more images and spent more time in each sitting than volunteers who only classified on
* Volunteers who contributed to both workflows still contributed many more yes/no classifications than survey classifications.
* There was greater inequality in volunteer contributions on the Yes/No workflow than on the survey workflow, though neither measure is outside the range of successful Zooniverse projects. The differences in Lorenz curve are consistent for both pre-existing users (those who had classified prior to the experiment) and new users (who joined Snapshots at Sea during the experiment).
* Accuracy – it’s actually a bit tricky to compare. I think all we care about is false negatives, because false positives get filtered out. If that’s the case, they’re pretty equivalent, though survey could have the tiniest of edges (need to .
  + After applying Ted’s aggregation rules for % agreement necessary, it looks like false negative rates are pretty equivalent at capturing the presence of a whale or dolphin (0% Type 2 error through both approaches), capturing the presence of a tail (5% yes/no and 2.2% survey), and capturing the presence of a humpback given a tail (0% false negatives for both). **However** false negatives on presence of tails carries over to the identification of a humpback in the yes/no approach, whereas it doesn’t necessarily in the survey approach; volunteers can correctly identify a humpback in an image even if they incorrectly identify that the underside of the tail is visible.
* Demographic/anecdotal survey responses (will probably never be able to include, given the turnaround)

**Discussion**

Key discussion points:

* Collaborative filtering provides an effective and efficient route to classifying an otherwise complex task.
  + While the simple yes/no questions engaged a similar number of volunteers as the survey workflow, each volunteer spent more time and contributed more classifications than on the survey workflow. Thus, despite the collaborative filtering approach being inherently less efficient, data were processed more quickly via this route.
  + Limitations: Snapshots at Sea is a relatively simple survey question, and the collaborative filtering approach only requires 4 distinct binary questions to achieve the same level of information as a survey question. Thus, a collaborative filtering approach may not be all that inefficient. In contrast, trying to use collaborative filtering to identify all 60 species from Snapshot Serengeti would require significantly more questions.
  + Classifications on the yes/no workflows are much faster on mobile. Also, despite no announcement of Snapshots at Sea’s recent availability on mobile phones, mobile classifications make up ~20% of all classifications contributed and > 30% of yes/no classifications, suggesting that mobile is a viable and largely unexplored route for contribution.
  + Need something about machine learning – ummm, are there actual existing techniques that could be applied here?
* Collaborative filtering does appear to engage volunteers differently than the survey task – the volunteer contribution was more unequal, regardless of whether volunteers had previously classified on SAS or not.
  + However, the implications of this inequality, and how much skew is “too much”, are not clear.
  + The collaborative filtering contributions are well within the range of other successful Zooniverse projects.
  + Users on the survey task only comment more frequently on the discussion forums than those who only classify on the yes/no tasks; however, these numbers are dwarfed by the engagement from users who contribute to both workflows. Since volunteers who contribute to both workflows still contribute overwhelmingly to yes/no questions, perhaps the difference in engagement reflects something about the interest of the volunteer or a limitation of the device on which they are classifying, instead of whether the task itself encourages engagement.
* Take-home message: collaborative filtering approaches work really well in this case – people like them and they get through data quickly. However, this type of workflow does appear to change the relative engagement of volunteers in the community, increasing the inequality of classifications and perhaps encouraging users who engage little on Talk. [I don’t think this is bad, I think it just might be an alternative route for users who don’t want to engage all that much].
* Applications: Perhaps as a filtering step in an otherwise fully complex survey task, say, filtering out empty images or those with the most common animals (e.g. filtering out wildebeest and blanks in SS would reduce the dataset by 80%). Allowing users to engage in multiple routes to classification would provide opportunities to engage with the images at different levels, according to the volunteer’s desires.