PROCEEDINGS OF SCIENCE

Highlights from Citizen Science in Astronomy and Planetary Science

Megan E. Schwamb[1]
Institute for Astronomy & Astrophysics, Academia Sinica
11F AS/NTU, National Taiwan University, 1 Roosevelt Rd., Sec. 4, Taipei 10617, Taiwan
E-mail: mschwamb@asiaa.sinica.edu.tw

PoS(BASH2015)001

BASH 2015
18-20 October 2015
The University of Texas at Austin, USA

[1]Speaker

1. Introduction

Known as citizen science, the process of hundreds of thousands of people directly contributing to research in the sciences and humanities has grown dramatically over the past decade. With the advent of the Internet, tens of thousands of people across the globe can be enlisted to help scientists with tasks that are impossible to automate or would be insurmountable for a single person or small group of individuals to undertake. The combined assessment of many non-expert human classifications can best the performance of a trained expert and in many cases outperform the capabilities of neural networks and automated routines [1][2][3][4][5][6]. As astronomical surveys and space-based mission data have grown with increasing data volumes (with the Sloan Digital Sky Survey [7], the Palomar Transient Factory [8], and NASA's Kepler mission [9][10] as just a few of the many examples) a citizen science approach has been applied to many of these datasets, becoming an established data reduction technique. In addition, projects have developed to not only use human pattern recognition for analyzing data but also for data collecting from tracking the impact of light population to the location of aurora. For a general overview of citizen science and web platforms for developing these online projects, we refer the reader to the Chapter by Trouille and the recent review by [11]. In this Chapter, we instead highlight a selection of active astronomy and planetary science focused online citizen science projects and consider the future of the field with the advent of the next generation astronomical facilities coming online in the next decade.

2. Highlighted Citizen Science Projects

2.1 Aurorasaurus (http://www.aurorasaurus.org)

Aurorasaurus aims to the study the physics and processes involved with aurorae on the Earth. The project is a website and application for reporting observations of aurorae and the observers' location. The project also gathers tweets referencing aurora and enlists the public to review the content from the Twitter stream and select those reports that are most relevant to the goals of the project. Through the geolocation information embedded in the tweets and the location information provided by the online reporting forms and application, the project creates a real time map of the extent and strength of aurorae. These observations are used to improve computer models in order make better future predictions about the behavior and extent of aurorae on the Earth.

2.2 Globe at Night (http://www.globeatnight.org)

The aim of the Globe at Night project is to study the impact and spread of light pollution through a crowdsourced network of observations of the night sky throughout a given year and spanning several years. The project collects measurements of the night sky brightness around the globe by members of the general public, using a web application (which includes a night vision version for dark adaption). Volunteer observers use the naked eye to assess what are the faintest stars that can be seen in known constellations while reporting the weather and sky conditions at the time of the observation.

2.3 Galaxy Zoo (http://www.galaxyzoo.org)

Galaxy Zoo is the first project from the Zooniverse, the largest collection of online astronomy citizen science projects to date. Going on nearly 10 years, this project aims to characterize the shape and characteristics of galaxies through human pattern recognition. Galaxy Zoo has examined data from the Sloan Digital Sky Survey, the Hubble Space Telescope, the Dark Energy Survey, and others. Over 40 peer-reviewed publications have been derived from Galaxy Zoo data, including the discovery of the first quasar ionization echo from a time when a now quiescent central super massive black hole was previously in an active AGN state [5]. For a more in depth overview of the project, I refer the reader to [4] and [12].

2.4 Asteroid Precovery (http://www.laeff.cab.inta-csic.es/projects/near/main/)

Volunteers are enlisted to identify known Near Earth Asteroids (NEAs) in ground-based telescope images to enable better orbit determination for potential impact hazards and to enable follow-up studies of the physical characteristics of the NEA population. Volunteers images that potentially have previously imaged the newly discovered NEA based on orbit predictions to better improve the know position and orbit of NEAs. For a more detailed overview of the project, I refer the reader [13].

2.5 Asteroid Mappers: Vesta (http://cosmoquest.org/?application=vesta_mappers)

The aim of this planetary science project is to study the surface of Vesta, the second largest body in the Solar System's asteroid belt. Volunteers from this CosmoQuest project are asked to review high-resolution images from the Dawn spacecraft's framing camera [14][15] and map the sizes of craters on Vesta, in addition to identifying boulders and bolder fields. The sizes and locations of the craters will be used to study Vesta's impact history as well as date the age of different surface regions on Vesta.

2.6 Research and Education Collaborative Occultation Network (RECON) (http://tnorecon.net/)

RECON aims to observe the stellar occultation of Kuiper belt objects (KBO), the remnants left after the construction of our Solar System's planets orbiting out beyond Neptune. With predictions of the timing and locations of the shadow path on the Earth, the amateur astronomers and public across the United States are involved in collecting observations to measure the exact time the star's light is blocked by the Kuiper belt object. With these observations, the size of the Kuiper belt object can be very accurately determined and also important for searching for atmospheres around the largest of these bodies.

2.7 Moon Mappers (http://cosmoquest.org/?application=simply_craters)

Moon Mappers [16] is a planetary science project from the CosmoQuest platform aimed at studying images from the Lunar Reconnaissance Orbiter Camera (LROC) Narrow Angle Camera (NAC). Volunteers review cutouts from the LROC NAC images to measure the locations and sizes of the sizes, shapes, and locations of the the craters will be used to study impact processes on the Moon and the collisional history of the Moon.

2.8 JunoCam ground-based observing and target selection effort (https://www.missionjuno.swri.edu/junocam)

JunoCam [17] is the optical and near-infrared camera aboard NASA's Juno mission scheduled to arrive at Jupiter in July 2016 when Juno will enter a polar orbit about the gas giant. Where JunoCam will observe on Jupiter will be decided with input from the general public. In particular, the JunoCam team will be working in collaboration with the amateur astronomer community to obtain full disk images of Jupiter. JunoCam itself will not be able to provide these types of images with Juno's polar orbit. The ground-based observations provide the much needed context for the cloud belts active storms on Jupiter to be able to determine where to image next with JunoCam. Members of the public will be able to provide input on which regions of Jupiter they would like to see photographed by JunoCam.

2.9 Planet Hunters (http://www.planethunters.org)

NASA's Kepler mission [9][10] has spent four years staring at a single patch of sky, simultaneously monitoring the same ~160,000 stars for these signatures of transiting exoplanets. Later as the K2 mission [18], the Kepler spacecraft is continuing to monitor fields along with the ecliptic for ~70 days with ~7000 stars monitored every 30 minutes. The Zooniverse's Planet Hunters project enlists the general public to review 30-day segments of Kepler observations to look for the characteristic dip in light due to a transiting exoplanet, and in particular to identify planet transits that may have been missed by automated algorithms in order to study the frequency and variety of bodies produced through different planet and solar system formation mechanisms. Planet Hunters has discovered over 30 planet candidates and several confirmed planets including an unusual transiting system that is potentially a transiting comet swarm informally known as 'Tabby's star'[19], and the first confirmed planet in a four-star system [20]. For a more detailed overview of the project, I refer the reader to [21].

2.10 Planet Four (http://www.planetfour.org)

During the spring on Mars' South Pole, carbon dioxide geysers loft dust and dirt through cracks in a thawing carbon dioxide ice sheet to the surface where it is believed that surface winds subsequently sculpt the material into dark fans observed from orbit. Understanding the direction, frequency, and appearance of these fans (a proxy for the geysers) and how varying factors impact these properties is important for understanding the Martian climate. With the Zooniverse's Planet Four project, volunteers help planetary scientists by mapping these dark seasonal fans visible in images taken by the HiRISE (High Resolution Imaging Experiment) camera [22] board NASA's Mars Reconnaissance Orbiter.

2.11 Radio Galaxy Zoo (http://radio.galaxyzoo.org)

The Zooniverse's Radio Galaxy Zoo [23] project aims match radio jets and outflows from super massive black holes located at the heart of galaxies and pinpoint the associated host galaxies in the infrared observations. With the extended nature of the radio data as well as the lower spatial resolution compared to optical and infrared observations, automating this task has proven difficult, but has proven successful with a citizen science approach.

PoS(BASH2015)001

2.12 Stardust@Home (http://stardustathome.ssl.berkeley.edu/)

NASA's Stardust mission collected grains from the Comet Wild 2 in its aerogel samplers, but in addition a rare number of interstellar grains passing through the Solar System were also serendipitously captured into the aerogel of Stardust Interstellar Dust Collector as well. The aim of Stardust@Home is to find interstellar grains by volunteers reviewing digital scanning electron microscope micrographs of the retrieved Stardust aerogel to spot perpendicular tracks and impact craters made as the grains entered from the direction perpendicular to the plain of the Solar System and became lodged in aerogel. The project has already identified seven particles that have properties consistent with an origin in the interstellar dust stream outside our Solar System [24].

3. The Future of Citizen Science and Citizen Science in the Petabyte Era

We are standing at the precipice of the petabyte age. The 2020s bring new challenges with the Large Synoptic Survey Telescope (LSST) [25], Square Kilometre Array (SKA) [26], and the end of ESA's Gaia mission [27][28]. ESA's Gaia mission is measuring the stellar parallax of nearly a 1 billion stars in the Milky Way over its 5-year nominal mission. When complete, LSST, an 8.4-meter optical telescope, will create the largest public dataset in the world, generating roughly 15 terabytes of data each night. LSST is expected to capture images of approximately 1010 galaxies and discover tens of thousands of Kuiper belt objects. With radio dishes to be constructed in Australia and in Southern Africa and combined to produce an effective collecting area of 1 square kilometer, the SKA will be the largest radio telescope ever built. The SKA will produce roughly 11 terabytes of raw radio data per second, more information than the combined traffic of today's Internet.

These never-before-seen data volumes are a technical challenge for both automated analysis routines and for citizen science. Even for a traditional citizen science approach the data volumes will too large to handle with current classification rates. Citizen science will be a way forward to tackle these never-before-seen data volumes, but it too will need to evolve to accommodate the coming data deluge. LSST will provide snapshots of billions of new galaxies, but it took 14 months for Galaxy Zoo 2 to classify 304,122 galaxy images [12]. The current citizen science model of choosing a random image/piece of data to be classified by a randomly selected volunteer just won't work in the petabyte era. The solution will be to combine crowdsourcing and machine learning algorithms as well as finding novel methods for efficiently using the time and energy of human classifiers. The bulk and routine tasks will have to be given to robot classifiers trained previously on human-based classifications. The citizen scientists will need to take on the harder and more specialized tasks as well as being enlisted to a review a subset of things the computers classify to continue to teach and improve the automated algorithms.

Additionally, ways of optimizing how many volunteers are needed to assess a specific image of piece of data and deciding images to serve to a specific classifier will be needed in this new era of big data. For most citizen science projects, images are randomly served to a classifier and data analysis performed after the fact. To get effective mine LSST and SKA data, there is

PoS(BASH2015)001

the need to implement smarter and more sophisticated algorithms to decide what data should be assigned to which classifiers to review. A first effort was with the bayesian Space Warps Analysis Pipeline (SWAP), which ran offline and determined based on volunteer performance when an image should stop being classified [29]. Now is the time build, test, and implement algorithms to direct volunteer effort in live time based and volunteer performance and real time assessment of the classifications. This will be an area of great focus before LSST has first light.

Identifying mechanisms to promote volunteer engagement will also be important in the next generation of citizen science projects. A recent analysis of Snapshot Serengeti, which enlists the public to review camera trap data and identify animals visible in the images, identified an anti-correlation with the time volunteers spent classifying and the number of scenery images with no animals. The more images without animals present, the longer the the volunteer engaged in the task and reviewed images on the Snapshot Serengeti website [30]. This has important implications for how future projects handling LSST-like datasets engage with the human classifiers. For example, optimizing for what only the machine classifiers can not handle, could leave only the hardest cases to citizen scientists to review. Like the behavior observed in Snapshot Serengeti, this may make them leave the site only after a few classifications. Citizen science projects in the petabyte era will need to also optimize not only for the capabilities of machines but also for human behavior. Task breaks and interventions may be other techniques for keeping volunteers engaged in online tasks and improves their retention and efficiency. Platforms like the Zooniverse and CosmoQuest would have the flexibility to provide volunteers breaks from their main/preferred tasks by presenting other projects and data in a similar topic or preference area for then volunteers for classifying. Further research in these areas are needed to identify the best method for maximizing human contribution for a citizen science approach when the SKA and LSST come online.

The prospects for citizen science in astronomy and planetary science are bright and will continue to be a tool in the arsenal of astronomers and planetary scientists to explore and tackle large datasets. In the past five years, the development of such websites and classification interfaces were built on a project by project basis where the bottleneck became the need for a dedicated web developer to build the front end classification interface and associated backend infrastructure and databases. Even with the rise of standard backend architectures to store volunteer classifications for these projects, the need for a developer to build the web tools has been a rate limiting step. This has changed in the past year, with the launch of the Zooniverse's new project builder system in June 2015. The Zooniverse project builder (http://www.zoooniverse.org/lab) enables researchers on their own to quickly and easily create a citizen science project from a standard set of web interface templates for annotations, decision trees, and sorting tasks that have repeated through. I anticipate that as as platforms like this continue to grow and develop, there will be a significant rise in the number of projects involving the public in the scientific process.

References

[1] Simpson RJ, Povich MS, Kendrew S, Lintott CJ, Bressert E, et al. 2012. MNRAS 424, 2442

[2] Khatib, F. 2011, Nat. Struct. Mol. Biol., 18, 1175

[3] Dashti, H. T., Ardalan, A., Siahpirani, A. F., et al. 2010, in Proc. of 2010 , International Conference on Artificial Intelligence, ed. H. Arabnia, D. de la , Fuente, E. Kozerenko et al. (Las Vegas, NY: CSREA Press), 183

[4] Lintott, C. J., Schawinski, K., Slosar, A., et al. 2008, MNRAS, 389, 1179

[5] Lintott CJ, Schawinski K, Keel W, van Arkel H, Bennert N, et al. 2009, MNRAS, 399,129

[6] Mendez, B. J. H. 2008, in ASP Conf. Ser. 389, EPO and a Changing World: Creating Linkages and Expanding Partnerships, ed. C. Garmany, M. G. Gibbs, & J. W. Moody (San Francisco, CA: ASP), 219

[7] York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, AJ, 120, 1579

[8] Law, N. M., et al. 2009, PASP, 121, 1395

[9] Koch, D. G., Borucki, W. J., Basri, G., et al. 2010, ApJ, 713, L79

[10] Borucki, W. J., Koch, D., Basri, G., et al. 2010, Science, 327, 977

[11] Marshall, P. J., Lintott, C. J., & Fletcher, L. N. 2015, ARA&A, 53, 247

[12] Willett, K. W., Lintott, C. J., Bamford, S. P., et al. 2013, MNRAS, 435, 2835

[13] Solano, E., Rodrigo, C., Pulido, R., Carry, B. 2014. Astron. Nachr. 335, 142-149.

[14] Russell, C. T., et al., 2012. Science 336, 684 - 686.

[15] Russell , C. T., et al., 2013. Planet Sci. 48, 2076 - 2089.

[16] Robbins, S. J., Antonenko, I., Kircho , M. R., et al. 2014,Icarus, 234, 10

[17] Hansen, C. J. Caplinger, M. A. Ingersoll, A., et al. 2014, SSRv., 39

[18] Howell, S. B., Sobeck, C., Haas, M., et al. 2014, PASP, 126, 398

[19] Boyajian, T. S., LaCourse, D. M., Rappaport, S. A., et al. 2016, MNRAS, 457, 3988

[20] Schwamb M. E., Orosz J.A. , Carter J. A. , et al. 2013, ApJ, 768, 127

[21] Schwamb, M. E., Lintott, C. J., Fischer, D. A., et al. 2012, ApJ, 754, 129

[22] McEwen, A. et al., 2007. J. Geophys. Res. 112

[23] Banfield, J. K., Wong, O. I., Willett, K. W., et al. 2015, MNRAS, 453, 2326

[24] Westphal AJ, Stroud RM, Bechtel HA, et al. 2014. Science, 345,786

[25] Ivezic Z, Tyson JA, Acosta E, Allsman R, Anderson SF, et al. 2008., ArXiv, e-prints

[26] Braun, R., Bourke, T., Green, J. A., Keane, E., Wagg, J., 2015, Advancing Astrophysics with the Square Kilometre Array (AASKA14), 174

[27] Perryman, M. A. C., de Boer, K. S., Gilmore, G., et al. 2001, A&A, 369, 339

[28] Lindegren, L., Babusiaux, C., Bailer-Jones, C., et al. 2008, in IAU Symposium, Vol. 248, IAU

[29] Marshall, P. J., Verma, A., More, A., et al. 2015, MNRAS, 455, 1171

[30] Bowyer, A, Maidel, V., Lintott, C., et al. Crowdsourcing, Proceedings of the First Association for the Advancement of Artificial Intelligence Conference on Human Computation (HCOMP 2015) Symposium