



Contribution ID: 43

Type: **Oral Presentation**

Citizen Science and Machine Learning: Towards a Robust Large-Scale Automatic Classification in Astronomy

Wednesday, 1 June 2022 16:20 (20 minutes)

Citizen science, traditionally known as the engagement of amateur participants in research, is demonstrating a great potential for large-scale processing of data. Using the power of the web, virtual communities of volunteers have been able to coordinate the classification of hundreds of thousands of images in a reasonable amount of time. In areas such as astronomy or geo-sciences, where emerging technologies generate huge volumes of data, this approach entails image classification at a rate not possible to accomplish by experts alone, although at the expense of worse quality in the classifications made by amateur participants. Despite its success in astronomy, as evidenced by the numerous editions of the Galaxy Zoo project, the current and upcoming massive surveys highlight its limitations, and the inclusion of machine learning methods towards a more robust automatic classification is considered mandatory. However, current efforts attempting the exploitation of citizen science outcomes with machine learning tools have ignored their inherent uncertainty as well as the potential of expert classifications to ameliorate this issue. Their ultimate goal has mainly been to replicate the amateur performance, thus propagating their biases and limitations and disregarding the fact that, apart from the data labelled by amateurs, there is also available (limited) expert knowledge of the problem along with vast amounts of unlabelled data that have not been exploited yet within a unified learning framework.

Our research delves into the development of automated approaches for astronomical classification problems that have been aided by citizen science projects on the web, aiming to leverage the inherent uncertainty in their results and all levels of knowledge available about the problem. We introduce an innovative learning paradigm for citizen science projects in astronomy capable of taking advantage of expert- and amateur-labelled images, and unlabelled images. As an implementation of this learning framework, we present the Citizen Science Learning (CzSL), an algorithm that first learns from unlabelled data with a convolutional autoencoder, and then exploits amateur and expert labels via the pre-training and fine-tuning of a convolutional neural network, respectively. As a case study, we focus on the classification of galaxy images from the first edition of the Galaxy Zoo project, from which we test binary, multi-class, and imbalanced classification scenarios, although the methodology is not limited to any classification problem in particular. Our results demonstrate an improved classification performance in comparison to a representative set of baseline approaches, showing a more comprehensive use of these resources that are available to the astronomical research community.

Presenter: JIMÉNEZ, Manuel (Instituto de Astrofísica de Andalucía (IAA-CSIC))

Session Classification: Outreach and Citizen Science