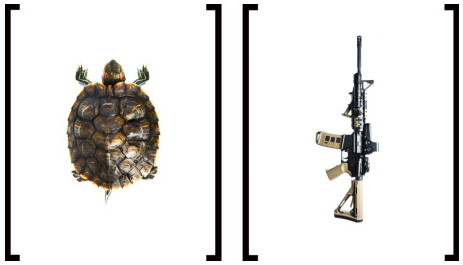
PROJECT REPORT CITIZEN SCIENCE

DENNIS KREUSSEL & TIM KAELBLE



CitizenAI

August 2020 –

ABSTRACT

Short summary of the contents in English... a great guide by Kent Beck how to write good abstracts can be found here:

https://plg.uwaterloo.ca/~migod/research/beckOOPSLA.html

ZUSAMMENFASSUNG

Kurze Zusammenfassung des Inhaltes in deutscher Sprache...

CONTENTS

1 INTRODUCTION 3

2 ROADMAP 5

2.1 Procedure based on Hackley's level of participation 5

2.1.1 1. Phase 5

2.1.2 2. Phase 5

2.1.3 3. Phase 5

2.1.4 4. Phase 5

2.2 Different Approaches of Analysing Citizen Science Projects 6

3 CHALLENGES 7

3.1 Privacy 7

3.2 Overlap with Community Science 7

1

INTRODUCTION

Citsci.cauterize is a citizen project about evaluating robust machine learning algorithms in different test scenarios. Researchers and citizen scientists alike can publish their algorithms and take part in a community endeavor to find general robust machine learning algorithms.

With regards to open source development and stackoverflow we see that technical experience and solving real world problems is always a valuable activity. Often these activities are used in CVs to certify problem solving skills. We will try to motivate people to contribute by giving them the opportunity to certify their experience in malware analysis via the web platform

- Scientist need humans as loss functions
- Loss function easy for recognizing images
- harder approach for sound, text, code, etc.
- integration of human knowledge
  → reinforce robustness of trained models
  → use human as reference

Since Machine Learning (ML) and other learning algorithms are used more widespread in various parts of science and in all kind of different enterprises, it is crucial for society to get a grip of what these algorithms are. With committees discussing in how far these algorithms, often referred to as „Artificial Intelligence “ or „AI “, should be allowed to interfere with peoples live, and the idea that the society has to decide how to handle with modern self learning algorithms it is decisive that the society or the citizens are being educated in that specific field.

This project is built on four main ideas. Firstly we want citizens to get in touch with different kinds of ways to use ML algorithms. Here we want to lower the boundary of accessing a field of computer science which people often refer to as intangible. With emphasis of people who may have never thought algorithms and computer science at all. Our goal in this first part of the project is to present an *explainable AI*. We want to show which features of an images lead the algorithm to its decision. With the interaction of the user on choosing if the algorithm was right with its decision we hope educate and to motivate all kinds of citizens to further investigate this interesting part of computer science. After the user had the opportunity to take the first steps in

this area we are confident that many will continue with the second phase of our project.

The second phase is the first part of citizens participation in science. Most of the ML algorithms that are being used need a feedback of how they are performing. For example if a robot wants to learn to jump it needs feedback of what a good jump is. Is a high jump a good jump? Is a proper landing important or more important than the height of the jump? What other ways of judging a jump exist? These and other questions should not be judged by people who develop the algorithm (who are they to decide such delicate subjects), but hopefully by a myriad of people to get a good feedback for the robot. Of course these decisions get even more difficult with AI applications in court, in police and other sensible areas. So not only training of the algorithms is necessary but also of the citizens. A more closer look in how this progression of feedback is actually benefiting these algorithms will be delivered in the following. This includes not only helping with speeding up the learning process but also hardening the algorithms.

The third part is a continuation of the idea of designing more hardened and robust algorithms. As in the open source approach safety and security of systems is more likely to be accomplished with a lot of reviews of the code rather than hiding the code and gaining security through obscurity with hiding the code as good as possible. This idea does not exclude ML algorithms. We want to provide a platform for developers to present there algorithms. We will also implement the possibility to upload data with which the algorithm may have difficulties or even judge wrongly.

With a collection of different approaches in the same problem category the algorithms can categorized by their robustness to the uploaded adversarial examples which will be presented in the fourth and last part of our project.

In this project we are trying to find an embedded approach of citizen science. We first want to educate to take a proper field of science closer to the people so in future it might be undoubtedly considered as citizen science.

2 ROADMAP

2.1 PROCEDURE BASED ON HACKLEY'S LEVEL OF PARTICIPATION

Participants can access contribution to the project in a step wise manner. Idea to iterate through Hackley's level of participation to reach more possible users. In an early phase a contributor is not obligated to have any prior knowledge about the subject. While helping with diverse Projects in the entry level, the contributor might feel comfortable with accessing more specialized tasks. In the following the different phases of contribution is discussed using *„Hackley's Level of Participation "*.

2.1.1 *1. Phase*

- Presentation of pictures
- Explanation for decision of algorithm
  → Is algorithm right with decision?
- Motivation for further investigation of algorithms
- Getting to know machine learning algorithms

In the first phase users are presented with images and a label based on the algorithms decision. The users are also be able to see a

2.1.2 *2. Phase*

- show different pictures/videos/data
- User decides which data fits the given label best
- User as loss function replacement/addition

2.1.3 *3. Phase*

- user can upload designed models
- other users can test models and upload possible attacks

2.1.4 *4. Phase*

- Ranking of robustness of different models for different pruposes

- collaborative analysis
- Forum for discussions

2.2 DIFFERENT APPROACHES OF ANALYSING CITIZEN SCIENCE PROJECTS

3 CHALLENGES

3.1 PRIVACY

- restrict user access
- possible malware is uploaded
- danger of tracing users actions?

3.2 OVERLAP WITH COMMUNITY SCIENCE

- Difficulties for users with no prior knowledge
- are we still in scope of citizen science