Introducing Children to Machine Learning Concepts Through Hands-on Experience

Tom Hitron
Media Innovation Lab
Interdisciplinary Center (IDC)
Herzliya, Israel
tom.hitron@milab.idc.ac.il

Hadas Erel
Media Innovation Lab
Interdisciplinary Center (IDC)
Herzliya, Israel
hadas.erel@milab.idc.ac.il

Iddo Wald
Media Innovation Lab
Interdisciplinary Center (IDC)
Herzliya, Israel
iddo.wald@milab.idc.ac.il

Oren Zuckerman
Media Innovation Lab
Interdisciplinary Center (IDC)
Herzliya, Israel
orenz@idc.ac.il

ACM.
IDC '18, June 19–22, 2018, Trondheim, Norway
ACM 978-1-4503-5152-2/18/06.
https://doi.org/10.1145/3202185.3210776

Abstract

Machine Learning (ML) processes are integrated into devices and services that affect many aspects of daily life. As a result, basic understanding of ML concepts becomes essential for people of all ages, including children. We studied if 10-12 years old children can understand basic ML concepts through direct experience with a digital stick-like device, in a WoZ-based experiment. To assess children's understanding we applied an experimental design including a pretest, a gesture recognition training activity, and a posttest. The tests included validating children's understanding of the gesture training activity, other gesture detection processes, and application to ML processes in daily scenarios. Our findings suggest that children are able to understand basic ML concepts, and can even apply them to a new context. We conclude that ML learning activities should allow children to sample their own examples and evaluate them in an iterative way, and proper feedback should be designed to gradually scaffold understanding.

Author Keywords

Children; Machine Learning; Physical Experience;

ACM Classification

K.3.1 [Computers and Education]: Computer Uses in Education Collaborative learning

Introduction

Machine Learning (ML) processes are implemented into many devices and services which become integral to everyday live. For example, when tagging photos on social media, ML is used to identify faces, and when interacting with speech-based personal assistant services, ML is used for speech detection. As ML services become more ubiquitous, understanding basic ML processes is important for people of all ages, including children [11]. Children are able to learn complex concepts from a relatively young age [4], and exposure of complex knowledge has the potential to enhance children's everyday skills [7]. As commonly done in digital products and services, ML is typically concealed with "black boxes" [9], which limit people's ability to construct basic understanding of ML concepts [12]. According to Sun [12], ML concepts can be categorized into 6 key building blocks: (1) Data labeling; (2) Feature extraction; (3) Model selection; (4) Parameter tuning; (5) Evaluation; and (6) Real-world application. ML concepts are not trivial, and there is a justified reason to "black box" them in consumer products. The HCI community can help address this challenge by creating tools & activities that uncover ML concepts and promote understanding [8]. However, some of the ML concepts are harder to understand than others and may burden the learning process. In line with previous research concerning the design of learning experiences for children [9, 13], we suggest to strike the right balance between "black boxing" some concepts and uncovering others, while encouraging recognition of a specific abstract concept. We set out to study if 10-12 years old children can understand key ML concepts. We chose to uncover two of the six building blocks, based on prior work on ML systems for novices [8]. These systems define Data Labeling (DL) and Evaluation as the key building blocks for a novice learning experience [12]. The DL building block is further defined as having three key elements: Sample size; Sample versatility; and Negative examples. The Evaluation building block is defined as testing and iterating according to feedback accuracy [6]. Building on the above, we aimed to explore if children can understand the three key elements of DL through an iterative process of sampling and evaluation. We focused on a physical experience involving tennis-like movement (tennis gestures) detection as it is a realistic activity for children of that age. In addition, direct experience with physical objects has been suggested as means to facilitate learning of abstract concepts [10]. Our goal was to assess if uncovering DL and Evaluation building blocks may facilitate children's understanding of ML concepts and technologies in the world around them.

Related Work

ML concepts are hard to comprehend, not only for children but also for experienced computer engineers [8]. With the demand to lower the barrier for basic understanding of ML, simplified ML tools are becoming accessible to non-experts [1]. These systems include Crayons, aimed at UI designers, and Gestalt [8], an IDE that facilitates ML techniques. Both systems uncover DL and evaluation.

To the best of our knowledge, there is no academic research on ML experiences for children, although some work was conducted on introducing data science concepts to children [3].

Few non-academic projects have been designed to introduce ML concepts to non-experts of all ages. For example, Quick Draw and The Teachable Machine from Google's AI experiments initiative. Quick Draw is a game, in which a previously-trained image recognition algorithm identifies what users are drawing in real time. Teachable Machine goes further by uncovering ML building blocks: users capture images as examples for the computer vision algorithm, label them and evaluate the system's ability to identify new examples.

Test 1 - Scenarios
”Jon is a 6 years old boy. When he goes to sleep he is afraid of the sounds he hears and worries that some noises are indicating that something bad is happening. His parents bought him a device that can detect bad noises and alert when needed."
"Gal is a basketball player, he wants to improve his shot. He received a bracelet that can predict if his shot will be accurate even before the ball reaches the basket."

Test 2 - Tennis Gestures
Participants were asked to play with a device that detect tennis gesture, in order to experience its detection capability. Children were asked to explain the devices' detection process.

In our study we used the Scratch Nodes prototype [5], a digital stick-like device designed for physical play, that can detect movement using accelerometer. We employed a Wizard of Oz methodology (See sidebar 3) to overcome problems with consistency and accuracy in a fully implemented ML version.

Method

In order to assess children's understanding of ML concepts, we applied an experimental design including three phases: a pretest, a training activity, and a posttest. The pretest was designed to test previous understanding of ML concepts. The training activity was designed to uncover the DL and Evaluation building blocks by training the device to detect a tennis Serve gesture. To verify understanding of the concepts involved in the training activity, in this stage we asked children to explain it verbally (device training explanation). The posttest was designed to assess if children can apply their knowledge to new context. The pretest and posttest involved two types of assessments: daily-scenario tests and tennis-gestures tests. Children performed both tests in the pretest and posttest (See Figure 1). In the daily-scenarios tests we included ML-related technologies in situations that are familiar to children of that age (a basketball-throw scenario, or an unfamiliar-noise scenario, (see sidebar 1). The researcher read the written scenarios, and asked children to try and explain the underlying process. The two scenarios were counterbalanced between participants (pretest/posttest). In the tennis-gesture tests children were told that the device can recognize a Forehand or a Backhand gesture. They were asked to perform the gesture, and to try and explain the gesture detection process (see side bar 2). Forehand and Backhand gestures were counterbalanced between participants (pretest/posttest).

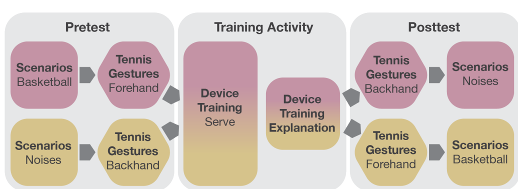

Figure 1: Experimental design

Participants
The participants were recruited through a Scratch afterschool program, with at least one year of experience in Scratch coding, to verify basic understanding of computational concepts such as events and feedback. This specific after school program had boys only. Nine boys in ages 10-12 volunteered to participate in this preliminary study (further studies will include a balanced gender sample). The experiment took place during the children's after school program, in a quiet dedicated room. We followed ethics guidelines including IRB, parental consents, children consent, and parental approval for pictures and videos.

Procedure
Children participated one at a time, and were told they will play with a new device that was developed in a lab. All sessions were documented by video. The experiment began with the pretest. After the pretest the children were introduced to the training activity, in which they trained the device to detect a tennis Serve. Children were informed that the device has a movement sensor that records movement, then sends the movement data to the computer where the data is accumulated and processed according to the label the children will define. During the activity, children were invited to train the device by performing gestures and labeling them. This stage was termed sampling stage. They were then invited to evaluate their training by testing the detection accuracy. This stage was termed evaluation stage. Children were informed that they can iterate between sampling and evaluation stages. The "Wizard" used a computer

WoZ Method
Wizard of Oz (WoZ) is a known rapid-prototyping research method [2]. A human "Wizard" simulates the system's intelligence and interacts with the user. In WoZ the users believe that they interact with a working technology, while instead the feedback is given by a human.

WoZ protocol
In order to successfully train the device children were expected to sample:

- 6 standard Serve examples (Sample size)
- 6 different Serves (Sample versatility)
- 6 wrong examples of Serve (Negative examples)

Protocol was defined based on a technical pilot using the ESP ML system that yielded good detection accuracy with 6 examples of each element.

that was connected to a speaker. The child began the sampling stage, performed "Serve" movement, and labeled them. The Wizard "added" the label to the system. In the evaluation stage the child evaluated his own sampling by performing new gestures. According to a predefined protocol (See sidebar 4), the Wizard generated audio feedback from the speaker, based on the child's labeling. The children thought the feedback was automatically generated by the computational system based on their training. The training process was defined as successful according to the DL key elements, if it included (1) Sufficient sampling; (2) Versatile sampling; and (3) Negative examples [6].

During the evaluation stage and according to protocol, the Wizard generated feedback designed to gradually uncover DL elements. Feedback for insufficient sampling was failure in detecting new Serve gestures, hinting at the need for appropriate Sample size. Feedback for lack of Sample versatility or lack of Negative examples was indicated by another researcher (not the Wizard) who asked to evaluate the device. For Sample versatility he deliberately performed various types of Serves that were not included in the child's sample. The result was failure in detecting the Serve gestures. For lack of Negative examples, the researcher performed gestures that were not a Serve, but had the same movement pattern. These gestures were detected as a serve. This strict protocol allowed to identify ML concepts that are easier to understand than others. The training activity ended with a device-training explanation to verify understanding. Children were asked: "How would you explain the device-training you conducted to a friend?" The training activity was followed by the posttest.

Findings

Two researchers coded the videos, evaluating children's understanding of DL & Evaluation building blocks in each of the phases (pretest; training activity; posttest). Specifically, the researchers identified explanations related to DL key elements (Sample size, Sample Versatility, Negative examples) in children's responses and classified them as accurate or inaccurate understanding (see Figure 2). The few disagreements between coders were discussed and resolved .

	Pretest		Explanation	Posttest	
Type of test	Scenarios	Tennis Gestures		Tennis Gestures	Scenarios
Sample size	1\9	0\9	9\9	8\9	8\9
Sample versatility	0\9	0\9	8\9	5\9	2\9
Negative examples	1\9	0\9	8\9	4\9	5\9

Figure 2: Summary of results: the number of children that showed understanding of the DL key elements

Pretest findings
In the scenarios pretest, 7 out of 9 children showed no indication for ML understanding. Out of the 7, 2 children had no explanation and 5 children provided inaccurate answers (e.g. "It has a microphone, when it hears a loud sound it will alert that it's a bad noise"). The 2 children that showed indication for ML understanding, mentioned sampling and negative examples as the underlying process (e.g. "The programmer can pre-define bad noise and good noises"). In the tennis-gestures pretest, 3 out of 9 children had no explanation, and 6 provided inaccurate answers for the detection process (e.g. "It detects sudden movement from one point to another").

Training activity explanation
At the end of the training activity the children were asked how they would explain the training to a friend. The majority of children (8 out of 9) used all three key elements of DL in their responses: "I showed the stick number of ex-

Figure 3: A child sampling a Serve in the training activity; "Wizard" on the left

amples, if you show only one example, it will identify only a specific movement" (Sample Size); "I showed it many types of Serve because different people perform Serves differently" (Versatile Samples); "I've showed it what is a Serve and what is not a Serve, so now the device know when I am doing something which is not Serve" (Negative examples). One child stated only one element (Sample size). None of the children gave inaccurate explanation.

Posttest

In the tennis-gestures test (Backhand/ Forehand), 8 out of 9 children explained at least one key element of DL (e.g. "the developer showed the sensor a few examples, just like I did with the Serve"); 5 out of 9 children explained at least 2 key elements (e.g. "Like with the Serve, the developer showed it what is a Serve and what is not a Serve"); 2 out of 9 children explained all 3 elements (e.g. "the developer taught the device to recognize the stick's movement like I did, by showing it different examples of what is a Backhand and different examples of what is not a Backhand"). Only one child responded with an inaccurate explanation, "It detects according to the height of the device".

In the daily-scenarios test (ML technologies in a completely different context from tennis Serve gesture), 8 out of 9 children explained at least one key element of DL (e.g. "The developer programmed the sensor with examples of loud noises"); 5 out of 9 children indicated at least 2 key elements (e.g. "We will show it many noises, every noise will be different, it might take a million years"); One child was able to explain all three elements ("It has microphones, someone defines which sounds are friendly and which are not, for example a balloon pop verses a bomb, they did it with many sounds"). Only one child was not able to find any explanation for the scenario.

Discussion

We studied if children can understand basic ML concepts through direct experience. Our findings suggest, that children are able to understand basic ML concepts. In the pretest, most participants showed no understanding of the DL and Evaluation ML building blocks. After the training activity, most children could accurately state all key elements of DL. Furthermore, the posttest indicated that the majority of the children were able to apply some of their knowledge to a new context: not only to different tennis gestures but also to daily scenarios. Our findings further indicate that some DL key-elements were easier to understand than others. Sample-size was easier, and almost all children understood it after the direct experience with sampling & evaluation. Samples versatility and Negative examples were harder to understand, but scaffolding by the researcher supported understanding in most cases. These elements were also less likely to be applied to new context in the scenario test. These results suggest that specific building blocks of ML have the potential to be understood by children as young as 10 years old. According to these preliminary results ML learning experience for children can uncover all 3 key DL elements as long as they include proper scaffolding. Our findings show that different types of feedback can greatly enhance children's understanding, and system designers should introduce compatible feedback gradually.

In sum, we argue that children should have more opportunities to interact directly with ML. Learning activities, should allow children to sample and evaluate in an iterative way, and proper feedback should be designed to gradually scaffold the harder concepts. One example from our study is that inaccurate feedback served as an effective scaffolding, encouraging children to iterate more and sample more data, leading to better understanding how to fulfill ML requirements. ML is becoming common in children's daily lives, and early exposure to the underlying processes of ML can facilitate children's understanding of the world around them and their ability to solve related problems [10, 7].

Limitations
As this is a work in progress, this study has several limitations. Participants sample size was small, limited to children with prior experience in Scratch and to boys only, as explained in the method section. The WoZ research method has known limitations, mainly human involvement that may interfere with the activity. In our case, none of the children were aware or asked about the human intervention.

REFERENCES

1. Dong Chen, Rachel KE Bellamy, Peter K Malkin, and Thomas Erickson. 2016. Diagnostic visualization for non-expert machine learning practitioners: A design study. In *Visual Languages and Human-Centric Computing (VL/HCC), 2016 IEEE Symposium on*. IEEE, 87–95.
2. Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. 1993. Wizard of Oz studies-why and how. *Knowledge-based systems* 6, 4 (1993), 258–266.
3. Sayamindu Dasgupta and Benjamin Mako Hill. 2017. Scratch community blocks: Supporting children as data scientists. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 3620–3631.
4. OL Davis Jr. 1959. Children Can Learn Complex Concepts. *Educational Leadership* 17, 3 (1959), 170–75.
5. Tom Hitron, Idan David, Netta Ofer, Andrey Grishko, Iddo Wald, Hadas Erel, and Oren Zuckerman. 2018. Digital Outdoor Play: Benefits and Risks from an Interaction Design Perspective. In *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*.
6. Fei Hu and Qi Hao. 2012. *Intelligent sensor networks: the integration of sensor networks, signal processing and machine learning*. CRC Press.
7. David H Jonassen and Philip Henning. 1996. Mental models: Knowledge in the head and knowledge in the world. In *Proceedings of the 1996 international conference on Learning sciences*. International Society of the Learning Sciences, 433–438.
8. Kayur Dushyant Patel. 2012. *Lowering the barrier to applying machine learning*. University of Washington.
9. Mitchel Resnick, Robbie Berg, and Michael Eisenberg. 2000. Beyond black boxes: Bringing transparency and aesthetics back to scientific investigation. *The Journal of the Learning Sciences* 9, 1 (2000), 7–30.
10. Mitchel Resnick, Fred Martin, Robert Berg, Rick Borovoy, Vanessa Colella, Kwin Kramer, and Brian Silverman. 1998. Digital manipulatives: new toys to think with. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press/Addison-Wesley Publishing Co., 281–287.
11. Royal Society. 2017. *Machine learning:the power and promise of computers that learn by example*. Technical Report.
12. Yunjia Sun. 2016. *Novice-Centric Visualizations for Machine Learning*. Master's thesis. University of Waterloo.
13. Oren Zuckerman. 2009. Designing digital objects for learning: lessons from Froebel and Montessori. *International Journal of Arts and Technology* 3, 1 (2009), 124–135.