The Impact of Visual Representations in a Machine Teaching Interaction

Claudel Rheault, Lorne Schell, Doriane Soulas, Chris Tyler, Tara Tressel, Masha Krol
6650 St-Urbain. Suite #500
Montreal, H2S 3G9
claudel.rheault@elementai.com

Abstract

Supervised machine learning typically requires labeled data examples to train models, and those examples can come as inputs from humans. By contrast, machine-teaching focuses on the teacher and its interactions with data, as well as its efficiency in teaching the model. Clustering is a visualization method that plots similar data points near each other. In this paper, we present a machine-teaching demo that can be run with a data clustering component that visualizes results and error rate performance of the model. Our main objective is to understand how a teacher can be influenced by how the model evolves and performs while training it. Specifically, we aim to understand how teacher perceives the capacity of a model to explain the root cause analysis of specific tasks, also known as explainability. Results from our preliminary pilot study with the demo suggest that there might not be any difference in perceived impact on the model with or without seeing live clustering, but motivation to do further teaching increases when seeing the clustering.

One of the main challenges for integrating AI into production in both the development and deployment phases is the lack of structured labeled data. In fact, in order to be relevant, current AI models require assistance of human operators, often acting as machine-teachers, to label data used as initial input (Sun et al. 2017; Sarkar 2016). Continuous cleaning and monitoring is then required to ensure the validity of data (Polyzotis et al 2017). It can require both expert and non-AI expert machine-teachers, depending on the task at hand (Sarkar 2016). For example, data labeling that requires complicated prior knowledge needs experts to teach a machine specific tasks (e.g. identifying nucleotides of DNA). Therefore, making the process of teaching easier and more accessible will potentially open up the production of machine learning solutions (Simard et al. 2017).

When approaching machine-teaching activities, a range of factors can play a role in the user experience. Human-centered design methods are central in the development of interfaces and systems that generate training data (Lindvall, Molin and Löwgren 2018). Previous work has shown that explainability, or the capacity of explaining to the teachers [the human operators] the rationale underlying of every action of a model [the learner] , can improve teaching effectiveness and general understanding of how the model works (Cakmak and Thomaz 2014; Jiang and Canny 2017). Explainability components can assist the user in interpreting the output of a machine learning model (Preece 2018).

Given this, we propose a machine-teaching demo to help understand how non-AI experts interact with machine-teaching labeling tasks. We explore the following questions in this paper:

1. What are non-AI expert machine teachers attitudes toward teaching machines?
2. Does model explainability have an impact on attitudes and motivation for future machine-teaching activities?

Related Work

Non-AI Expert Teachers

Machine teaching interfaces for non-AI experts have been explored by industry leaders in AI development. Microsoft has built a platform for interactive concept learning, where users can give features to a model and provide labels to build a model (Microsoft 2018). Google has a teachable interface where the user uploads images and then teaches the machine what is contained in the image (Google 2017).

Active Machine Learning and Machine-Teaching

Active learning is a method that provides the learning algorithm control over the overall learning process by selecting the most important unlabeled dataset items for human annotation (Cohn, Ghahramani and Jordan 1996). Building from this notion, extensive research has been conducted in developing active machine-teaching interfaces (Cakmak and Thomaz 2011; Johns et al. 2015). In this paradigm, the user has to perform a larger set of tasks. For example, the user can be involved in the selection of which data to train the model on, thus optimizing the dataset. Previous work has also investigated prototypes that can allow the user to judge on the model's result and optimize the training process (Jiang and Canny 2017; Sun et al. 2017).

Presented in the Work in Progress and Demo track, HCOMP 2019.

Figure 1: Machine-Teaching Interface Demo

Machine-Teaching Demo

Rationale

The purpose of this demo is to prototype a machine teaching interface, and test the hypothesis of clustering to provide transparency on how the model is learning:

- Do users want to see how and if they can influence the model?
- Does seeing the impact on model encourage the user to continue?
- Do users value the machine teaching activity?
- Do users that have knowledge in AI value machine teaching tasks more?

System and Interaction

The demo we present in figure 1 is a machine-teaching interface combined with a live data visualization of what the model has learned. We use Active Learning (Kirsch, van Amersfoort and Gal 2019) to reduce the number of training samples from MNIST (LeCun and Cortes 1998) required to train the model. We create the visualization using t-SNE (van der Maaten and Hinton 2008) clustering on the last learned feature layer of the deep learning model.

The interface is divided in three main areas: the teaching interaction, the clustering and a graph of the error rate. The user is expected to indicate which numbers is displayed on the image, and the explainability components are being updated with the user input, with every 5 samples collected.

Preliminary Study

Pilot Study

To test the initial assumptions stated earlier, we conducted a preliminary test with a sample group. The test included a pre-survey, a machine-teaching activity and a post-activity survey. One moderator per participant was present in order to guide the user between the different steps. The pre-survey was used to assess the general level of understanding of AI, and the user's interest to do machine-teaching.

In the machine-teaching session, participants had to correctly identify the number shown on the image by selecting the corresponding number on the buttons. The interface had 2 versions: Version A showed the full demo as explained above, and version B showed only the image with the buttons, with no visualization component. The post-activity survey asked the user to reflect on his potential impact on the model, and to reassess future level of motivation to do machine-teaching activities. A total of 24 users participated in the test. 10 were shown version A, 14 version B

Limitations

Testing environment was not controlled, as it was in an open event. The sampling was limited to the crowd attending the event. The length of the test was restricted due to event schedule, and time-on-task discounted as a measure due to limited participant time.

Preliminary Observations

Both groups have a high interest in teaching a machine prior to the test. A majority of participants (82,9%) believe humans have an important role in teaching AI. Self-reported data suggest that participants who engaged with the visualization of clustering might be more open to do machine-teaching exercises again, and for a longer period of time. As it is now, the perceived impact of machine-teaching on the model and the understanding of what was taught is similar across the two groups.

Conclusion & Future Work

Machine-teaching performed by non AI-expert user will play a growing role in how we build and train models. The interface on which it is performed is the main interactive element that can provide feedback and Intel to the user, having a potential impact on the machine-teaching interaction. Our demo adds visualization to the training activity, and this can potentially play a role in the machine-teaching interaction.

This demo will be used to continue investigation further with a broader sample of participants and a more detailed study of the impact of showing model's evolution to the teachers. As providing instruction and guidance as to how the machine performs and can prove to have a positive impact on overall experience of teaching (Cakmak and Thomaz 2014), future work could integrate the notions of context and instruction to guide the teaching experience. We also aim to investigate potential expectations of the user on machine-teaching. Humans tend to assume that AI is better than what it is (Sethumadhavan 2019), thus inducing a form of expectation of the machine-teaching activity. Disappointment and frustration can then take place when the systems isn't matching the user's expectations (Dosilovic, Brcic and Hlupic 2018). In upcoming research, a better assessment of user's expectation and understanding of AI would need to be performed before engaging in the machine-teaching activity.

References

Andreas Kirsch, Joost van Amersfoort, Yarin Gal, 2019. BatchBALD: Efficient and Diverse Batch Acquisition for Deep Bayesian Active Learning*arXiv:1906.08158*

Cakmak, & Thomaz. 2014. Eliciting good teaching from humans for machine learners. *Artificial Intelligence*, 217, 198-215.

Cakmak, M., & Thomaz, A. L. 2011. Mixed-initiative active learning

Dosilovic, F., Brcic, M., & Hlupic, N. 2018. Explainable artificial intelligence: A survey. *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 0210-0215.

D. A. Cohn, Z. Ghahramani, M. I. Jordan 1996. Active Learning with Statistical Models. *Journal of Artificial Intelligence Research, Vol 4*, 129-145

Google Creative Lab, 2017. Teachable Machine. *https://experiments.withgoogle.com/teachable-machine*

Jiang, B. and Canny, J., 2017. Interactive machine learning via a gpu-accelerated toolkit. *In Proceedings of the 22nd International Conference on Intelligent User Interfaces* (pp. 535-546). ACM.

Johns, E. M., Aodha, O. J., & Brostow, G. 2015. Becoming the Expert - Interactive Multi-Class Machine Teaching. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 07-12, 2616-2624.

Laurens van der Maaten and Geoffrey Hinton, 2008. Visualizing Data using t-SNE. *Journal of Machine Learning Research* 9 (2008) 2579-2605

Lindvall, M., Molin, J., & Löwgren, J. 2018. From machine learning to machine teaching: The importance of UX. *Interactions*, 25(6), 52-57.

Microsoft Research, 2018. Machine Teaching Demo. *https://www.microsoft.com/en-us/research/video/machine-teaching-demo/*

Polyzotis, N., Roy, S., Whang, S., & Zinkevich, M. 2017. Data Management Challenges in Production Machine Learning. *Proceedings of the 2017 ACM International Conference on Management of Data*, 127746, 1723-1726.

Preece, A. 2018. Asking 'Why' in AI: Explainability of intelligent systems – perspectives and challenges. *Intelligent Systems in Accounting, Finance and Management*, 25(2), 63-72.

Sarkar, A. 2016. Constructivist Design for Interactive Machine Learning. *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 07-12, 1467-1475.

Sethumadhavan, A. 2019. Trust in Artificial Intelligence. *Ergonomics in Design: The Quarterly of Human Factors Applications*, 27(2), 34

Simard, Patrice Y., et al. 2017. Machine teaching: A new paradigm for building machine learning systems. *arXiv:1707.06742*

Sun, Y., Lank, E., & Terry, M. 2017. Label-and-Learn: Visualizing the Likelihood of Machine Learning Classifier's Success During Data Labeling. *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, 523-534.

Yann LeCun and Corinna Cortes, 1998. The MNIST database of handwritten digits. *http://yann.lecun.com/exdb/mnist/*