Human-In-The-Loop Machine Learning with Intelligent Multimodal Interfaces

Syed Z. Arshad SYED.Z.ARSHAD@IEEE.ORG
DATA61, CSIRO, 13 Garden Street, Eveleigh, NSW 2015, Australia

Jianlong Zhou JIANLONG.ZHOU@DATA61.CSIRO.AU
DATA61, CSIRO, 13 Garden Street, Eveleigh, NSW 2015, Australia

Shlomo Berkovsky SHLOMO.BERKOVSKY@DATA61.CSIRO.AU
DATA61, CSIRO, 13 Garden Street, Eveleigh, NSW 2015, Australia

Fang Chen FANG.CHEN@DATA61.CSIRO.AU
DATA61, CSIRO, 13 Garden Street, Eveleigh, NSW 2015, Australia

Abstract

In this paper, we argue that intelligent multimodal interfaces add an important dimension for advancing the cause of human-in-the-loop machine learning (HITL-ML). Multimodal interfaces seek to leverage natural human capabilities to communicate via speech, gesture, touch etc. Such interfaces are said to be intelligent when they can better learn and adapt to the requirements and condition of a user. Here we show how this implicit learning of system parameters (e.g. via interaction feedback loop) and labelling of user cognitive states is an effective and often overlooked dimension of HITL-ML. We also present a research brief of relevant investigations undertaken in this regard.

1. Introduction

The need for human-in-the-loop machine learning (HITL-ML) is now being popularly discussed not only in technical circles (Biewald, 2015) but also in business world (Bridgwater, 2016). It is generally thought that Machine Learning (ML) based systems can achieve about 80% accuracy at most (on their own), but this is not acceptable for most sensitive or mission critical systems. Improvement beyond 80% would require human interaction/intervention or assistance in some form. As per Lukas Biewald (co-founder and CEO CrowdFlower), human-in-the-loop machine learning can help in solving both training and accuracy issues (Gutierrez, 2016). Humans, not only help create training data by explicit labelling, but can also handle the tough judgement calls (e.g. understanding bad handwriting or parsing slang in language). This results in increased accuracy as these difficult judgments can then be used to further train a machine learning algorithm so that it can start handling more complex judgments. In fact, the tough and ambiguous examples are strongly recommended as training data.

Proceedings of the ICML2017 Workshop: Human in the Loop Machine Learning, Sydney, Australia, 2017.

However, interestingly, there is another dimension to this explicit user interaction; and that is the physiological signals and behavioural data of the user, as captured by intelligent multimodal interfaces (see Figure 1). The physiological signals input could be users galvanic skin response (GSR), blood volume pulse (BVP), pupil dilation, speech frequencies etc. whereas behavioural data could be from eye gaze, handwriting strokes (on tablet input) or user mouse movement patterns. This type of data has been used to successfully label cognitive load states of the user (Chen et al., 2016), making decision making measurable (Zhou et al., 2015) and also machine learning useable (Zhou et al., 2016). As depicted in Figure 1, user interaction can broadly be categorized into two components, namely (a) the conscious user input meant for the ML-based system and (b) the user behavioural and physiological data as captured by intelligent multimodal interfaces. In the rest of this paper, we will focus on the implicit HITL-ML (resulting from interaction data) but first a few words about intelligent multimodal interfaces and its significance.

1.1. Intelligent Multimodal Interfaces

When people interact with one another in a natural environment - they do so in a multimodal manner, making use of up to all five senses (in serial or parallel) to better understand a situation. Unfortunately, traditional computer

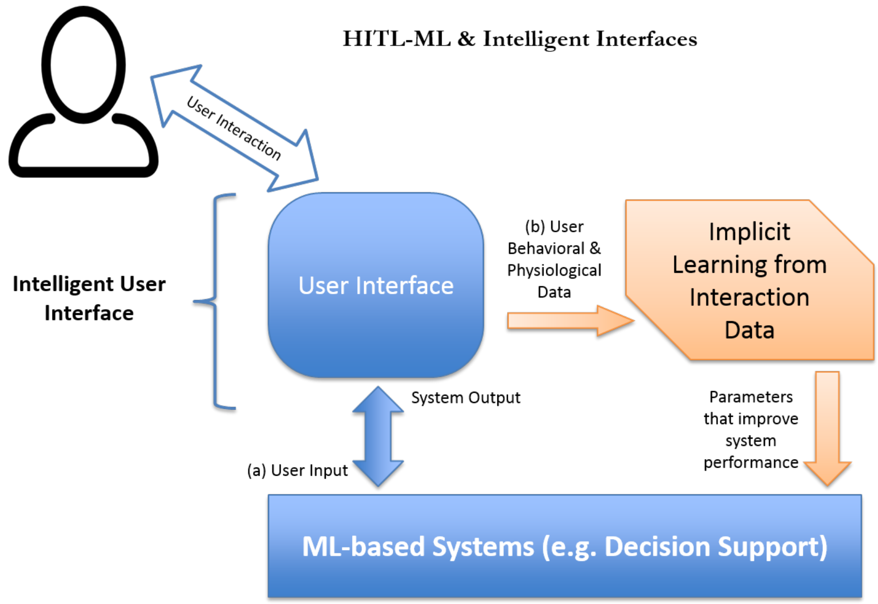

Figure 1. Human in the loop ML and intelligent user interfaces.

interfaces have relied only on fixed modalities/channels to communicate with its users. Historically, this has resulted in the instruction, practice and training of users to interact with interfaces in a manner that suited the systems processing capabilities. Luckily, all this is now changing with increasing popularity of human centred designs that advocate modelling users natural behaviour (including any constraints on users ability to attend, learn, and perform). The resulting interfaces are more intuitive, easier to learn and lesser prone to performance errors. Multimodal interfaces can be described as interactive systems that seek to leverage natural human capabilities to communicate via speech, gesture, touch, facial expression, and other modalities. This brings sophisticated pattern recognition and classification methods to Human-Computer Interaction (HCI), as multimodal interfaces need to be inherently flexible due to the requirement of mobility and universal access (Oviatt, 2002). More recently, due to significant advances in wearable and bio-sensor technologies, multimodal interfaces have started to directly make use of the physiological and behavioural signals emanating from various user modalities. The goal of research in multimodal interaction has been to develop technologies, interaction methods, and interfaces that remove existing constraints and move towards the full use of human communication and interaction capabilities (Turk, 2014).

2. Implicit Learning from Interaction Data

Several successful attempts have been made to further research in learning from interaction data. Here we choose to present one such concrete theme, its area of application and an ongoing research activity.

Multimodal interfaces are said to be intelligent when they can better understand and adapt to the requirements and condition of the user. A research theme that revolves around one such goal is multimodal cognitive load measures framework (Chen et al., 2013). Very briefly, the framework, when applied to a user interface, has the potential to monitor several of the available behavioural and physiological signals in order to infer (or label tentatively) the users cognitive load state. Once the state/label is confirmed, control feedback is provided to the system so that it can adapt accordingly. Here we would like to highlight the *feedback loop* formed when the system learns/infers about user (from user activity data streams) and then user reacts to (or labels) the decisions taken by the system (or tentatively displayed on screen). More details of the framework can be found here (Chen et al., 2016). With the advent of interactive machine learning we see a wonderful opportunity in terms of advanced techniques (at the level of learning from physiological and behavioural signals) for making multimodal interfaces even more effective. We believe that multimodal interfaces are a great area of application for interactive machine learning.

The idea of IML is not new to HCI. Earlier efforts includ-

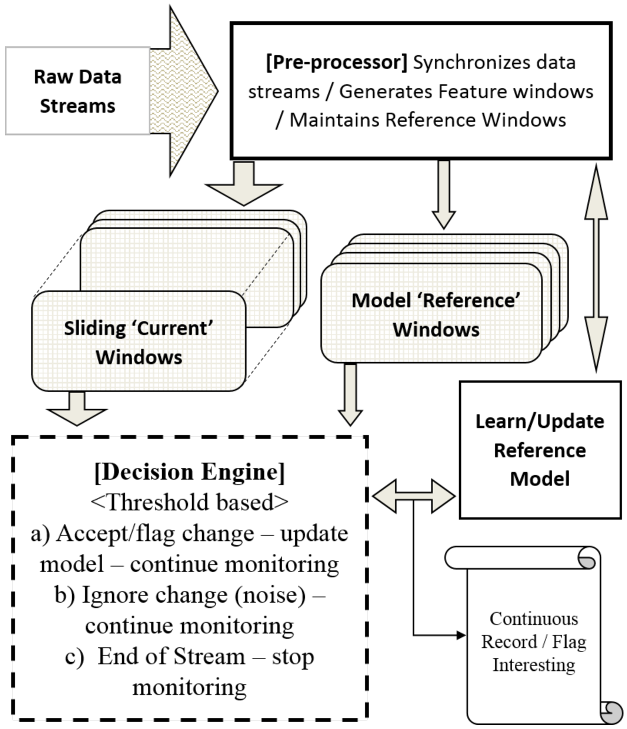

Figure 2. Real time implementation to detect user's cognitive load.

ed perceptual user interfaces like the classical Crayons tool (Fails & Olsen, 2003) that contrasted IML with classical ML model and assisted practically in image processing. More recently, the interaction loop of IML has been argued as a facilitator of constructivist learning, as it maximises the interaction between the end-users experience of the model, and their ideas regarding the model status (Sarkar, 2016). Constructivism is the view that learning occurs when ideas and experiences interact. In case of intelligent multimodal interfaces, this learning can be likened to a deeper understanding (or self-discovery) of ones own performance abilities and limitations through ones own physiological and behavioural signals.

There exist algorithms that learn behaviours via human feedback. Recently Loftin et al. (Loftin et al., 2014) proposed two algorithms that interpret human feedback as discrete communication depending on the behaviour the trainer is trying to teach and the teaching strategy used by the trainer. Teaching strategies considered in this research were reward or punishment focussed, balanced or inactive. They argued that as some human trainers use a lack of feedback, to indicate whether actions are correct or incorrect, the system may continue to learn by interpreting lack of feedback as implicit feedback.

Using basic multimodal cognitive measures framework, Arshad et al. (Arshad et al., 2015) formulated the problem of cognitive load detection from multimodal data as the problem of detecting concept drift from data streams. Their solution implemented sliding windows technique to detect users cognitive load in live scenarios. This implementation (see Figure 2) maintained separate (current and reference) windows for each learned feature. The decision engine was responsible for pooling in various feature change detections and then using a threshold to predict overall behaviour change (i.e. flag or ignore). This learning could in principle be either validated, rejected or ignored by the user. (Much of the learning here was implicit as the system was only designed to intervene when user cognitive load level was critically high). In case a load/behaviour change was validated, relevant feature reference model windows were updated using data from both current windows and freshly streamed points. This dynamic reference model update enabled the implementation to remain relevant while continuously monitoring for changes. Every successful change detected inspired a new norm for updated model. This implementation monitored eight user mouse behaviour features and the feedback loop was used. However, there remains room for much improvement and perhaps applying algorithm techniques proposed by (Loftin et al., 2014) could improve the results in such scenarios.

3. Challenges and Opportunities

From the framework details and the case of implementation briefly stated above, we strongly believe that interactive machine learning can greatly benefit making multimodal interfaces intelligent. However, learning from users physiological and behavioural streams of data is not without challenges. Firstly, it is quite complex to identify precisely the users cognitive state based on signals emanating from users behaviour of physiology. There could be multiple overlapping cognitive states at any given time. However, the research so far aims for the state that would result in observable behaviour.

Also, that learning with or without feedback (explicit or implicit) can conveniently take place for intelligent multimodal interfaces as well. However, there will be situations where lack of feedback cannot be ignored (e.g. in high performance or mission critical scenarios) and the system must take corrective measures and in extreme case forcefully override manual control.

The case of IML for intelligent multimodal interfaces can at some higher level be argued as learning about and then optimizing personal performance from ones own body physiology and behaviour as monitored and mirrored by the system. Several issues come to mind immediately. Why should the user trust the systems judgement/recommendation? Does the user not know better through proprioception? And the answer is simply that since the system continues to be interactively trained with users explicit or implicit feedback, the case is closer to that of a personalized classifier than a general one. Strong cases of power to the people and role of humans in IML (Amershi et al., 2014) have already been presented, but here we turn the focus of learning on humans themselves (as they perform in various environments) using their own physiological and behavioural signals. Research and development in this area could result in intelligent interactive personal coach like system interfaces that keep humans performing optimally over prolonged periods of time.

4. Conclusion

We conclude that intelligent user interfaces have a unique role to play in HITL-ML and that multimodal interfaces can greatly benefit from the techniques of IML especially in the domain of real time user performance management. In this paper, we highlighted an implicit dimension of HITL-ML and presented a derived implementation that demonstrates opportunity for using IML. Also, that learning from user physiological and behavioural signals to infer and validate human cognitive states can prove to be a unique and interesting area for interactive learning machines.

References

Amershi, Saleema, Cakmak, Maya, Knox, W. Bradley, and Kulesza, Todd. Power to the people: The role of humans in interactive machine learning. *AI Magazine*, 35(4): 105–120, 2014.

Arshad, Syed, Wang, Yang, and Chen, Fang. Interactive Mouse Stream As Real-Time Indicator of User's Cognitive Load. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 1025–1030, 2015.

Biewald, Lukas. Why human-in-the-loop computing is the future of machine learning. `http://www.computerworld.com/article/3004013/robotics/why-human-in-the-loop-computing-is-the-future-of-machine-learning.html`, 2015. [Online; accessed 4-June-2017].

Bridgwater, Adrian. Machine learning needs a human-in-the-loop. `https://www.forbes.com/sites/adrianbridgwater/2016/03/07/machine-learning-needs-a-human-in-the-loop/#2bbb379f4cab`, 2016. [Online; accessed 4-June-2017].

Chen, Fang, Ruiz, Natalie, Choi, Eric, Epps, Julien, Khawaja, M. Asif, Taib, Ronnie, Yin, Bo, and Wang, Yang. Multimodal behavior and interaction as indicators of cognitive load. *ACM Transactions on Interactive Intelligent Systems*, 2(4):22:1–22:36, January 2013. ISSN 2160-6455.

Chen, Fang, Zhou, Jianlong, Wang, Yang, Yu, Kun, Arshad, Syed Z., Khawaji, Ahmad, and Conway, Dan. *Robust Multimodal Cognitive Load Measurement*. Human-Computer Interaction Series. Springer, 2016.

Fails, Jerry Alan and Olsen, Jr., Dan R. Interactive machine learning. In *Proceedings of IUI2003*, IUI '03, pp. 39–45, 2003.

Gutierrez, Daniel. Human-in-the-loop is the future of machine learning. `https://insidebigdata.com/2016/01/11/human-in-the-loop-is-the-future-of-machine-learning/`, 2016. [Online; accessed 4-June-2017].

Loftin, Robert Tyler, MacGlashan, James, Peng, Bei, Taylor, Matthew E., Littman, Michael L., Huang, Jeff, and Roberts, David L. A strategy-aware technique for learning behaviors from discrete human feedback. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, pp. 937–943, Québec, Canada, 2014.

Oviatt, Sharon L. Breaking the robustness barrier: Recent progress on the design of robust multimodal systems. *Advances in Computers*, 56:305–341, 2002.

Sarkar, Advait. Constructivist design for interactive machine learning. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 1467–1475, 2016.

Turk, Matthew. Multimodal interaction: A review. *Pattern Recognition Letters*, 36:189–195, 2014.

Zhou, Jianlong, Sun, Jinjun, Chen, Fang, Wang, Yang, Taib, Ronnie, Khawaji, Ahmad, and Li, Zhidong. Measurable Decision Making with GSR and Pupillary Analysis for Intelligent User Interface. *ACM Transactions on Computer-Human Interaction*, 21(6):33, 2015.

Zhou, Jianlong, Khawaja, M. Asif, Li, Zhidong, Sun, Jinjun, Wang, Yang, and Chen, Fang. Making Machine Learning Useable by Revealing Internal States Update A Transparent Approach. *International Journal of Computational Science and Engineering*, 13(4):378–389, 2016.