Secret Lives of Machines

Ray LC

Parsons School of Design / Cornell Medical College and Cornell Tech

rayLC@newschool.edu

http://www.raylc.org/machines/machines.html

Abstract

Machines have taken over activities traditionally ascribed to human workers, from supermarket checkouts and vehicle driving, to analyses of our habits and taking over our habits. The future will require those working at the boundaries of self, machine, perception, and behavior. They have to know not only what makes us humans, but what makes machines machines, and they'll step inside the machine to see what they see as a way of understanding other minds. This requires a paradigm shift in our relationship with networks of machines. We have to experiment and play with a future where networks of machines are a part of us, for machines interacting amongst themselves serve as a template for how they interact with us. We created two sets of agents following different rules of behavior. One set follow human faces using a stationary camera on the back that keeps track of view locations. When humans move out of view, the group of agents do their own thing like chatting and dancing, but when humans come into view, their locations are updated and the agents all look directly at them in 3D alignment. Another set of agents follow a single camera mounted on one agent which tracks only human face that it can see. The viewer has to actively engage the camera before the group of agents can follow. Here, the machine doesn't have to build a model for where the user is; it moves towards viewers if faces are found. This model-based vs. model-free modes of interaction generate audience thoughts of flakiness, comfort, and irritability. We can play with properties that make these robots human-like. Using a network of machines that all respond to humans by computer vision, we examined their behaviors amongst themselves and behaviors synchronized in to us. Autonomous machine units demonstrate how we step inside to interpret machine intentions, creating embodied loops of human-machine-human interactions.

Fig 1. "Secret Lives of Machines" by Ray LC: installation sculptures imbued with active and passive interactive elements via computer vision and staged performance. Source: Ray LC**.**

Parents would always ask their children, "what do you want to be when you grow up?" When I was a teenager, I loved answering half facetiously "what *you* want me to be." Thinking back with the wisdom of a half grown adult, I realize that hidden within that answer unwittingly are two major thrusts on how to think about design.

Answering "what you want me to be" emphasizes the process of becoming. We never truly become who we want to be, because once we become that, we immediately strive for something else. In science we study the objectified world: we want to know how an organ works or what a chemical consists of. The fundamental departure in good design thinking is to go beyond the object and experience the process (Mauer, et al, 2017), to take as the primitive element not a thing, a state, or a goal, but rather a relation between things, a process amongst things. Thus as designers we think beyond constraints of the physical universe in a world ripe with possibilities and imagination, a world that is not measurably scientific.

The other nuance that comes with answering "what you want me to be" is that it acknowledges the need to agree with another person. What we desire each other to be is intimately tied to how we can empathize and communicate with each other. To make the world a better place, we can start by understanding each other, understanding ourselves.

In the digital age, human communication by itself is not enough to account for the diverse forms of understanding that we have to navigate. Just as our ancestors looked into nature and wondered at the existence of spiritual beings beyond themselves, we are now at a point in history where the objects and systems we create are doing things beyond ourselves that we no longer grasp entirely. Take the use of i2 EIA, an IBM Watson program, to identify terrorists posing as refugees (Stereyl, 2016). It claims to provide a score giving the probability that an asylum seeker is who she claims to be. Border officers don't know how it operates, and engineers answer with the vague notion that it "develops a comprehensive understanding of your threat landscape." These metrics are only one of a list of credit scores, academic scores, and online rankings that determine our image, including credibility, likeability, and productivity. While machines are abstracting data from our entire activity trace, we humans are more alienated from an understanding of how machines behave than ever.

Thinking back to the question of "what we want to be," this paper proposes to examine what machines have to become in the future as humans and machines form new styles of interaction. The decades old, anthropomorphic interactions of humans asking computers questions embodied in the Turing Test is too simplistic (Bratton, 2015).

In the face of an increasingly connected world of intelligent devices using information about what we do to infer who we are, we humans too, must look at machines and artificial agents in our own way and infer who *they* are. While machines are hardwired to quantify data, we humans are emotional beings who react to our peers. Thus in this future we experience how machines make us feel, how a group of machines interact amongst themselves in a network the way human lovers make sacrifices for each other, or how human mothers take care of their children, asking them who, or perhaps "what" they want to be.

Results

Interactive modular network for HCI

We experiment with the natural world because we want to see how it works, so why should we not also experiment with machines, and in particular a group of machines that interact amongst themselves? For example, we can present happy faces and sarcastically happy faces to different facial recognizers and ask how machines can tell the difference, whether we can fool some machines but not others. Then we can look at the properties of these machines to figure out what parameter they tuned that allowed them to figure out whether we are sarcastic or not despite only getting graphical information.

In this project, I decided to build robots which look the same but have different styles of interaction. We have two groups of agents following different rules of behavior initially. One set follow human faces using a stationary camera on the back of the platform that keeps track of view locations. When the user moves out of view, the group of agents perform their own activities such as chatting and dancing, but when the view comes into view, their locations are updated and the agents all look directly at them in alignment in 3D. Another set of agents follow a single camera that is mounted on one agent which tracks only the human face that it can see. In this case, the viewer has to actively engage the camera before the group of agents can follow her. The machine in this case does not have to build a model for where the user is; it simply moves towards the viewer if a face is found.

This distinction between the two groups of machines is the classic model-based vs. model-free AI algorithms found in artificial intelligence research (Russell & Norvig, 2009). The former constructs a model predicting where human faces are found and responds quickly and decisively when one is found, but it cannot generalize well in new situations where the terrains are different for the calculations. The latter keeps track only of a way of moving towards faces, a heuristic that moves itself to the right when a face is detected in the left side of the pixels, etc. It cannot detect all faces in the environment, but it can generalize to any situation where faces around, even outside the platform.

Video documentation: https://youtu.be/b8liAWU8XXM

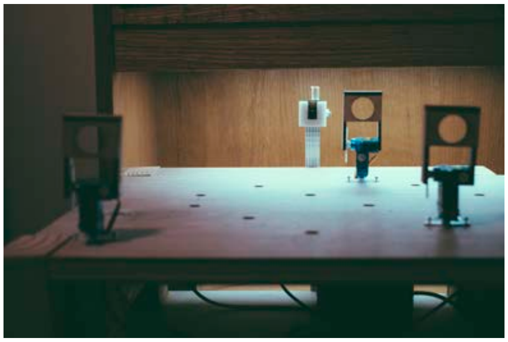

Fig 2. The model-based system constructs a model of the world where the human faces exist, by doing inference using a camera mounted at the back. This group of agents then move with instruction from the back camera. Model-based learning gives robust responses but at the mercy of inflexible designs and lack of hidden motives in the machine system. Source: Ray LC.

Fig 3. The model-free system uses a camera mounted on one of the machines that only follows the audience when her face can be detected from the machine's current position and rotation. This makes the robot much more believable as a being with human-like intelligence. Model-free systems are more general but at the risk of losing global knowledge of all positions of the audience. Source: Ray LC.

Intelligent modular network in human interaction

When audiences interacted with the system, they noticed that the model-based system can track their face fairly reliably, albeit after a slight delay. They often call it "cute but creepy," but when I turned a few of the agents of the model-based system into ones that do their own tasks without tracking faces, viewers find it much more natural and non-threatening, much like a "cute flock of owls." Although users had the most trouble with the model-free system (the agent with the camera attached and its associated followers), surprisingly they found it the most temperamental and "human-like." Perhaps it is because that system forces the user to move their face into the agent's field of view in order to be tracked, so the users are forced to "get the robot's attention," much as "prima-donna" type personalities in the real world. This mixture of a "self-centered" model-free system and a "reliable" model-based system give the users the richest experience where members of the flock appear to show emotionality to those interacting with them.

We can program the systems to perform in response to human interaction. For example, I produced a "Romeo and Juliet" script in which "Alice" and "Bob" agents shyly stand next to each other "in publc" when their compatriots are tracking viewer faces. When Viewer faces are not found, the other agents do their own activities, and "Alice" and "Bob" are able to "meet" (i.e. turn towards each other) and consummate their love. It took a few attempts for viewers to understand this, but once they did, they found the story amusing, and began assigning emotionality to behaviors of the flock that depend on their interactions with us. The act of seeing has engaged a new dynamic in the network where characters alter their connections.

We can see our reactions to these different styles and play with the properties these robots that make them human-like. Thus using a network of machines that all respond to humans by using computer vision, we can examine their behaviors amongst themselves as well as behaviors synchronized in some way to us. The machines serve as a metaphor for networks of interaction. Humans can begin to see these relationships amongst machines: flocking behavior reminiscent of birds, internal conflict of fractions within the group, the rebellious maverick of the group, or even love, care, or desire? Autonomous units with all range of movements allow us to demonstrate how machine intentions are interpreted by human viewers.

Fig 4. Interaction with audiences detected via computer vision determines behavior of the network of machines. If humans are watching, the machines are guarded with each other and perform a face-following after the audience. When audiences are not watching, the micromachines perform their own story, in this case the "Romeo and Juliet" story from Shakespeare. See video in the section referenced below for the narrative. Source: Ray LC.

Conclusion

In childhood, our parents asked "what do you want to become?" It is a metaphor for the current state of our society. Will we synchronize with the things we make? Will we react against machines to become more biological than ever? Who we want to be, or what we want to be, will be the central question in the loops of the future where we look at things we make and we look at ourselves. Lost amongst this dynamic is how things we make look at other things we make. This requires looking at networks of machines and how they interact with humans, and moreover, how their own interactions become ways we relate to them. All this converges to a view of the world where relationships and processes are the primitives for living, as opposed to objects and goals. We live in a networked world. The relationships and the process are the fundamental experiences in design, and relationships of networks of machine entities allow us to imagine in our own terms.