Enabling Technologies to Rethink Factory Automation

Ross A. Knepper

Most factory robots today look and work the same as they did thirty years ago. The design has been perfected by companies such as KUKA, ABB, and Fanuc over the years. What, then, are the factors limiting the more extensive adoption of automation in the factory setting? The interaction design of these robots is what we could call time-delayed teleoperation. A highly-skilled programmer instructs the robot in great detail how to perform a task. Then, the robot repeats the task later in the absence of humans. We frequently see such robots caged off from humans. These cages protect people from the robots, which are typically strong, fast, and unpredictable to anyone besides the programmer. The time-shifted teleoperation paradigm has reached the limit on its ability to innovate in the manufacture of goods, and a new interactive approach is now needed.

Beyond safety, the cage serves an additional function that is less well known, which is to protect robots from humans. These robots understand their tasks at only the most basic level. They do not comprehend how steps of a task relate to one another in space and time to accomplish larger goals. Thus, if humans disturb a workpiece or leave behind some clutter, the robot is unable to compensate by adapting to the change. Such interference instead triggers a failure that shuts down the system until a trained expert can come and fix it. Thus, humans can only be involved in the assembly task by taking turns with a robot.

If we want to advance the state of the art in automated factories, then robots need a higher-level semantic understanding of what they are programmed to do. If successful, this change will unlock a chain of further transformations in factory automation. If robots understand their task, then they will be robust to the presence of humans in their workspace, enabling humans to work more closely with robots. By moving away from time-shifted teleoperation, robot programs can be adjusted almost in real time, in response to failures or bugs in programming. This new approach to programming robots opens up many kinds of manufacturing to partial, human-supervised automation. Many new markets would exploit this technology, including smaller companies, products with small runs or frequently-changing processes, and products that are too complex for their production to be fully automated.

My research focuses on developing robot skills for higher level reasoning. For example, my IkeaBot system performs autonomous, multi-robot furniture assembly with decentralized computation. The user provides the system with CAD files of furniture parts, and IkeaBot determines how the parts might fit together (like a person who chooses to ignore the Ikea assembly manual and figure out the process on their own). The IkeaBot system then divides the tasks among the available robots and proceeds to assemble the furniture. The benefits of this system are that it is insensitive to changes in source part locations, it can adapt automatically to changes in the form of the parts, and it can even recover automatically from some types of failure.

One shortcoming of the IkeaBot system is that programming it is still complex and tedious. In my lab, we are building technologies to allow untrained users to interface productively with robots. These intuitive interfaces require robots to understand how people think and act. For example, we added to IkeaBot a new capability to ask a human for help when failures occur. In order to maximize productivity and minimize cognitive load, the robots should not express the problem to a human, but should instead offer a preferred solution. This reduces the cognitive load on the human user and reduces the need for situational awareness, which a user might not have if they are supervising several robots.

To ask for help clearly and unambiguously (or indeed to generate any communicative signal), the key insight is that the sender seeks to maximize not their own confidence that the meaning is correctly encoded in the signal, but the confidence of the listener that the meaning is correctly understood. Thus, our *inverse semantics* framework encapsulates as a core component of language generation a model of natural language understanding. The system generates several candidate help requests and tries to understand each one, comparing its understood meaning with the original intent. In selecting the shortest unambiguous form of the request, the generated sentence pithily incorporates context in the form of physical surroundings as well as commonly-understood circumstances.

More broadly, if robots are to be programmed by people who may not be experts in programming robots, then robots will need to learn the natural human "protocols" that we use in communication with one another. In contrast to the practice of professional software engineers, who are trained to completely specify all aspects of an algorithm, studies in psychology suggest that humans naturally prefer to underspecify a notion initially and then work out the details through interaction. This human practice saves substantial effort because we are able to reason out many of the unspecified details for ourselves. Robots must therefore also learn to extract unspecified details from the context in which programs are given.

The primary challenges to achieving this level of contextual understanding are related to perception and modeling. Robots need substantial semantic knowledge in order to correctly categorize and associate information they learn. The process is further hindered by the fact that robot sensors are noisy and error-prone, lending uncertainty to all readings and requiring probabilistic inference about complex domain-specific knowledge.

A promising avenue to help with perceptual understanding is deep learning, which has proven adept at visual recognition and classification tasks. Recently, it has also beaten the human "Go" champion and automated a financial office in Japan. The major challenge for deep learning in robotics is data. Deep learning is a part of the big data revolution, and it thrives in domains that can offer (for example) one million annotated training examples. In robotics, data collection often requires robot action, thus limiting the rate at which meaningful data can be gathered. Perhaps the next big advancement will be a "small data" movement that augments machine learning algorithms with symbolic or semantic understanding of a domain in order to shrink the size of training sets required.

In closing, it is worth reflecting on our goal in building new technologies for factory automation. We seek not to replace human workers with machines but to augment them and increase productivity. In fact, robots would not function without people in such complex automation domains, where minor failures and programming changes will be frequent. Consequently, robots and humans will increasingly need one another.