Mapping Autonomy

Leon Kester

Applications for Autonomy

Autonomous system play an ever increasing role in modern society. Typical applications where ethical issue play a role are mobility and defense.

For mobility autonomous vehicles may have to make decisions on how much risk to take, how close to drive to preceding vehicles or who's life to spare and who's life to put in danger. In the defense domain much attention goes to the problem of autonomous vehicles as weapon systems making decision on life and death.

In order to address these issues for various domains a generic approach is taken. First we map autonomy in automation and learning and then look at the role of the human. Then we look at a third aspect, next to automation and learning, i.e. self-adaptation and reconsider the role of the human. Finally considerations on ethics of such autonomous systems are discussed.

Autonomy: Automation

The basis for autonomy is the automation of situation assessment and situation management.

Automatic situation assessment may consist of sensing with various dedicated sensors, processing the signals of theses sensors, detecting interesting objects, following those objects in time, classifying the objects, identifying the objects, finding relations between the objects and determining the possible risks or benefits those object may cause. For objects that have a form of intelligence also the intent of the objects is important.

The development of automatic situation assessment is typically from the lower levels, i.e. signal processing, detection and following the objects towards the higher levels. Representations of the situation on the higher levels becomes more and more complicated and therefore more and more of a challenge.

Also at the higher levels the representations need to be more in consensus with the semantic representations humans use to characterize the situation. For the application domains like mobility and defense also the situations are not only complex but highly dynamic so the range of possible situations the autonomous system needs to assess is huge.

Automatic situation management uses the results of situation assessment, also called situation awareness to manage the situation. Automatic situation management has the same challenges as situation assessment, i.e. it becomes more difficult at the higher levels due to the larger complexity of the situation which results in complex representations. Algorithms are more tedious to develop and use more processing power and memory.

Another trend in automation is the cooperation between multiple autonomous systems with situation assessment and/or situation management capabilities.

Cooperation is easier at the higher levels of situation assessment and situation management since not much data is needed to exchanged such that the cost of communication is acceptable. However,

cooperation at the lower levels will result in more consistent and higher quality situation assessment and management. Since communication costs are getting lower the trend is distributed situation assessment and situation management at lower levels.

Advantages of automation is that it is relatively fast and transparent, meaning that the working of the algorithms can be analyzed in order to understand its behavior.

Disadvantage is that it is not dealing well with failures of the system and unexpected situations.

Learning

Another aspect of autonomy is learning. Learning autonomous systems are basically trying to assess and manage the situation. In case of a positive results this behavior gets a higher weight, otherwise a lower weight.

In learning online learning (learning on the job) and offline learning (training) can be distinguished.

There are many different flavors of learning, e.g. evolutionary algorithms, reinforcement learning, ant colony optimization, deep learning etc. It is beyond the scope of this paper to discuss the developments and advantages and disadvantages of these methods individually.

In general developments in learning are particularly in making learning faster and therefore suitable for reaching more complex goals in more complex situations. Learning will be faster because computers and communication becomes faster but also there is development in more efficient algorithms.

The advantage of learning is that it can be a powerful tool in case the situation is too complex to model completely.

There are a number of disadvantages, however. For real time applications it is still very slow. It is not transparent, meaning that is difficult to understand why and how it works. Online learning can be unpredictable and is prone to deceiving. Offline learning has the same disadvantage as automation in that it does not deal well with failures of the system and unexpected situations in which it had no training.

Role of the Human

With the developments of automation and learning in mind we can distinguish three different roles of humans with respect to (semi)-autonomous systems.

The human cooperates with the autonomous system. Both play their part in situation assessment and situation management. Since autonomous systems have still difficulties in doing situation assessment and situation management at the higher levels that is usually the task of the human. Next to the situation assessment and situation management capability of the autonomous system there is also a need from the human to the autonomous system to explain how the autonomous system works. So a matter of trust. Since automated and learning systems have difficulty in explaining how they work, cooperation between human and autonomous systems is a challenge.

The human can also act as a fail-safety agent. He intervenes when there are failures in the autonomous system or the autonomous system is in an unexpected situation such that erroneous behavior is expected.

The human can also act as a moral agent. The autonomous system is not expected to be able or is not allowed to make moral decisions. The human makes moral decisions or intervenes when the autonomous system is expected to behave immorally.

Self-adaptation

An aspect of autonomy that is often not addresses together with automation and learning is self-adaptation.

Self-adaptation first of all involves self-assessment. In what kind of situation am I? What is my goal given this situation? What is the current state of my resources: sensors, computation, communication, actuators? How well do I perform situation assessment and situation management? How useful is my information to others? How useful would information of others be for me? How do I appear to the outside world? How certain am I of what I am doing?

Then based on this self-assessment the autonomous systems can next to managing the environment manage itself. It has a possibility to reason: now that I know and understand myself better and my environment how should I manage myself and my environment in order to reach my goals?

Role of the Human revisited

When we also consider that autonomous systems may adapt themselves to changing goals, changing situations and changing own capabilities also the perspective of the role of the human changes.

If the self is also assessed and managed failures of the system can be recognized and managed. The role of the human is no longer the manager of all parts of the autonomous system but a task allocation similar to that for situation assessment and situation management is called for.

Also the cooperation between human and autonomous system is improved because the autonomous system is now able to explain what it is doing or trying to do while the human can express to the autonomous systems its needs. These needs become the goals of the autonomous system and can adapt its behavior.

This also open up possibilities for autonomous systems to take part in moral reasoning. Because not all specifications of the autonomous system are determined during the design it can derive from much higher level goals what would be desirable to do in this particular situation.

Human and autonomous system thus now also have a ground to cooperate in moral reasoning.

Consideration on Ethics of Autonomy

The goals that autonomous systems pursue, and therefore also the ethical goals, should be consistent and explicit. If there is a possibility that the goals cannot be reached also the usefulness (utility) of less desirable outcomes should be quantified. In this case the autonomous system can always try to get as close to the desired goal.

It is often argued that the goal (or utility) function is difficult to specify in this way. There are several reasons why this might be less difficult than presumed. If an autonomous system has a self-assessment and self-management capability it can derive utility during operation from the high level goal function so it does not have to be completely specified during design. There is a reluctance to be too specific about ethical goals so the difficulty of specifying the utility functions is sometimes used as an excuse for not doing it.

From a perspective of mission effective human autonomous teams, not only the current shortcomings and limitations of autonomous systems should be considered but also the shortcomings and limitations of humans, such as cognitive constraints, cognitive biases and cultural differences.

Much attention is going to autonomous systems that can be lethal or make decisions that may bring people at risk. However, that does not mean that developments of autonomous systems in other domains are not of concern. We should also in a more generic way address the control of autonomy itself.