Machine Learning,ML

Prof. Carl Gustaf Jansson
Prof. Henrik Boström
Prof. Fredrik Kilander
Department of Computer Science and Engineering
KTH Royal Institute of Technology, Sweden

Lecture 3
Intelligent Autonomous Systems and Artificial Intelligence

So this is the third lecture of the first week of the course on in machine learning and the topic of this lecture is intelligent autonomous systems and artificial intelligence. When artificial intelligence was founded in 1956, it was not so long ago less than 10 years since Thomas Watson CEO IBM, made a famous statement that 'probably not more than five computers are needed in the whole world'. Of course some time had gone since that day, but still we had a world where computers were rather rare, compare to other phenomenon, to other things. We may be have five computers but we didn't have thousands of computers, we didn't have millions of documents, have billions of computers and of course many early works in artificial intelligence could not avoid to be influenced that we talked about we talked about computers and we talked about whether the computers could do stupid things or intelligent things. Today the world has changed and of course also everything has to change with it. So today we don't talk about five computers we talk about billions of artifacts in the world that can be more or less intelligent and it can be vehicles on our roads, it can be vehicles in a terrain, it can be vehicles on the water, it can be like drones in the air, it could be all kinds of robots either in our industries but also in our homes like vacuum cleaners and other machines and we can also look at swarms of robots that can solve different problems. So we face now a situation where we are surrounded by potentially more and more intelligent autonomous entities and of course there is a role now for various scientific areas like artificial intelligence to try to make sense out of the situation we face. So what I want to discuss with you now is an attempt to have an abstract view of an intelligent autonomous system and for the moment let's forget that we have this gigantic system of billions billions of artifacts that all stinger' each one of them can be intelligent and of course what do we say about the whole system, that's another problem. So let's stay with looking at one of these entities at a time which

of course is a simplification but we have to start somewhere. So for any intelligent agent, human or animal or artificial, there are 13 things that must be in place. So first of all as an entity you have to have the model of the world you live in or act in, so this means that you have to have a representation of what surrounding you so the model of the world is crucial and it cannot have to have many forms but it have to be there, because if you don't have any model of the world how can you how can you perceive, how can you think how can you can't so this means the model must be in place. Given a model of the world you have the first process which is the perception, so when you act as an entity in this world which of you have a model of,you have the problem of efficient perception, which means that you have to have measurements make measurements or sensory observations and from these sensory observations you have to collect the relevant data in every situation and from the relevant data you have to create abstract and then given that you have a perceptions of the situation at hand then you need some reflection some analysis some reflection some reasoning about the situation as a background for actions because for an entity in the world you have to act there's no idea to perceive if you don't have an intention of acting. So this cycle of perceiving, thinking and acting is very central to any intelligent being and only steps and there are sub steps so for acting if you have to plan what to do and when you know have a plan you have to configure actions and when you have set up a good configuration of actions as with your plans then you have to perform this action. So we can say that and all these of course all these steps have to be performed in terms of a model of the world and then at the bottom here we have learning so this means of course for any intelligent being you cannot just repeat the same behavior time after time you perceive, you think, you act, but you have also then perceive again and observe the consequence of your actions and of course in order to improve you have to have some learning mechanism that affects your thinking and your actions and based on earlier experience, so this is one way of looking at how an intelligent system have to behave. So when we compare computer science and artificial intelligence what can each of these areas contribute with respect to building artificial intelligence systems, so of course you can do a lot with the computer science but what I want to argue here is there are some fundamental differences between what characterized the core methodology and technologies in computer science and the keywords that characterized computer science are keywords like determinism, causality, certainty, completeness, invariance and quantitative data knowledge. So the idea in computer science is that you can characterize the situation you are in, in a complete way and then you can

create algorithms that in a deterministic and causal fashion derives some consequences. So it's an ideal world in a sense because it presuppose at least for its partial situation some complete and certain information in order to arrive the inference and as we all know an intelligent autonomous system, lives that acts in the real world cannot always rely on such a situation, so this means that for many purposes computer science make sense because certain situations are complete and we can use computer science to devise methods that can make a system function however there are many situations which are inserted where we lack of complete information, and we cannot rely on only quantity data, we have to combine to use a qualitative data, quantitative data we may look need to look for abnormal deterministic solutions and we may also need to have a behavior of the system that is adaptive and not invariant. So for me this distinction these distinctions given these keywords are the crucial difference and for me it's the basic argument why the artificial intelligence techniques that set out to handle the non complete the uncertain, the dominant aesthetic mixed qualitative situation and in the long run will have a better prognosis of making useful systems. So if we now turn to how work research in artificial intelligence has been performed during the last six years, the key division work meet on the three aspects knowledge representation, automated reasoning and machine learning. Knowledge representation in focusing after what in the world, automated reasoning focused on how based on perception make the necessary steps to generate adequate actions and then finally machine learning which are the mechanisms to make it possible to improve the reasoning process, so this is one dimension then it's obviously so that for knowledge representation, artificial intelligence has worked with two kinds of representation types, symbolic and sub symbolic and this is have of course also effect the kind of reasoning you can perform so for symbolic representation it's one kind of reasoning to sub symbolic another, for symbolic representation is one kind of learning techniques and for sub symbolic another and you can see on this slides examples of the specific representation and the specific reason in a specific very needed, so this slides shows you how work in artificial intelligence can be indexed scientifically. But it reasonably well maps also on the earlier slide that came a picture how our intelligent tunnel system could function. The core artificial intelligence I tried to describe on the earlier slide, however artificially there is a border line of areas around core artificial intelligence which are many case assumed to be involved in the concept and I will call them the second rim here, an example of such areas are machine perception, computer vision, robotics. Most people would say of course robotics is artificial

intelligence but still a very specific area with the specific issues but sharing the core issues of artificial attempt. Another areas game-playing optimization is use in technical systems, dedicated data mining, language engineering, speech technology, intelligent interfaces our expert systems. These are all kinds of special systems that in most cases and in the name the conception of the area is counted as part of artificial intelligence but it makes sense to differentiate between these extensions from the core technologies and most of these extensions really share the essential parts of the core. So with all these advances that now happens regarding artificial intelligence of course there are many theoretical and philosophical issues that came to one's mine, so are theoretical limits for the intelligence of artificial systems and in that case what restrictions and other issue is of course that worries people a lot is a very practical issue will the emergence of intelligent efficient system dramatically reduce the human workforce. Obviously of the quite obvious and many of these intelligent artifacts meant to do tasks then it historically has been done by humans so obviously the existence of intelligent artificial systems will change the workload there will be more jobs or a few jobs remains to be seen but it's not so much a technical problem as a society problem another issue is will the excesses of the system dehumanize the human life which means as so many things happen are without human intervention so will our life get less human because we interact more with artificial agents, on the next level one can also then start to think are there some intention coupled to this kind of system or it has just do simple tasks so could one regard them as benevolent or malevolent, does it make sense to attribute ethics or morale to the artificial systems which of course then very closely relate is it meaningful to talk about that is these kind of systems of intentions, their consciousness, do they have a mind and of course the ultimate questions that is thought out is it so that at some point these systems will entirely take over so are the existence of the system an existential threat for the human race, it's called a singularity. I don't think that the present moment there is any clear answer to any of these questions, for the moment many of the applications are pretty simple so we are pretty much in the beginning of this development so therefore many of these worries is not very relevant initially but this development will go on and after a while many of these issues will come up and become more and more crucial over time. So thank you for your attention the next lecture will be now on applications of machine learning.