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 18
Inductive Learning based on Symbolic Representations and Weak Theories

Welcome to the fourth week of the course in machine learning this week as the title Inductive Learning based on Symbolic Representations and Weak theories. So first I want to make some comments on the title of this week, so inductive learning is kind of self-evident, so why is it called learning in symbolic representations. So I hope you have already understood there is for a long time been a balance in artificial intelligence between work based on symbolic representations and work on sub symbolic representations. And by symbolic representations is meant Symbols, Lists, Semantic Networks, Bayesian networks, Production rules, Logic decision trees many of those things that we introduced last week. In contrast to that we have Sub-Symbolic Representations Neural Networks, Genetic Algorithms, Binary Strings and you also heard something about that already. And for any part of artificial intelligence oh you could look over the years the focus of work has been on one or the other, and it has gone a little up and down. At the moment in machine learning there is a very strong focus on sub symbolic representation in particular Neural Networks but that does not mean that that all the other work such as all the other streams of work are dead or drying up and all the different approaches continue but at any point in time the line 90 days is one of them. What is important to say is that independent if you have a focus or one of the other symbolic or not sub symbolic there is also a common denominator here you know casein that mathematics because both can really describe what they do more or less purely in mathematics and mathematics is the foundation for both these approaches. So anything regarding groups, numbers and sets and graph factors lattice, tensor so to say and these are relevant under all circumstances, but essentially this week we will still focus on symbolic representations.
So let me know that is try to explain the reason for the last part of the title learning in Weak Theories and the motivation for including that part in the title is that there are two clearly two categories of work in machine learning. So I would say the largest amount of work has

been put on what is termed here learning in the presence of weak theories, in the meaning that this is creating an abstractions from sets of instances with a minimal very or very weak or almost absent model or backgrounds. Of course there is always some background knowledge I think we already talked about that for many of the methods and an algorithm where we are studying there are a lot of parameters that have to be set, we have talked about biases that can be included into the languages we use and so on and of course these all these things these biases permit these things constitutes background knowledge and we cannot be never be entirely without it but in many of the cases we set up a bar and many cases will be studied this week there is very very little such knowledge. So this means that we more or less entirely learn the abstractions from by generalization from the instances but typically also in order for this to work we need normally pretty large sets of instances. So that's the scenario we are focusing for the mode and the opposite situation which we will go more into coming week is how you can create abstractions but also more if are abstractions on the border of an existing model or background theory that is more substantial, it could be that is almost complete and just to be a adapted slightly or it can always partial and need to be extended. In these cases you may have a lot of instances but these techniques could also work with smaller set of instances because there is so made much guiding knowledge and for actually guiding the way the abstractions are formed so this is the reason for this distinction so this week we will still focus on induction in a Weak Theory scenario.

I will now give you a very brief overview of what's going to happen this week, so we will have four lectures apart from the last tutorial lecture concerning the assignments and these four lectures has the following teams. So the first is Generalization as search which is a classical abstract framework for this describing how generalization can be performed. Secondly we will look at learning algorithms for Decision Trees the further which was decision trees was introduced last week as a formula, now we will look into some algorithms in that area. After that we will turn to Instance Based learning and as mentioned on an earlier lecture already instance based learning is the approach where you actually not explicitly store or build up any abstractions rather that the abstractions are implicitly defined by the way we structure our memory of instances. In all these three cases we primarily focused on supervised learning which means that the instances we look at are all labelled typically with a class label, and also we also partially looked at the case where we have a numerical output but which is what we call regression, but still it's super vast. In contrast to that we will finish the week with a lecture on what is called clustering which is mean one of the technique type of techniques that is crucial for the situation where we have unclassified input and the situation which we here on the course have termed unsupervised learning. So in the first of these lectures generalization is certain we will introduce a very general framework introduced by Tom Mitchell many years ago, and one can say this is an interesting framework because in a very abstract way it tried to illustrate what happens when linear in early generalization takes place doesn't matter whether it takes place in one formulas more in the another. And we will spend some time on looking how what representation we can use and discuss that and then we will spend most of the time of what is termed data-driven strategies for generalizations which is essentially where we start from our instances and we actually based on a walk through the breadth-first walk through all the instances but to find in order to find the solution make a search of hypotheses space that we built up and, then we can do that either in a depth first manner or a breadth-first manner or in a kind of combined approach we call as the Version space. So this will be the first lecture. So next lecture will be about decision trees we will study a specific category here which we called TDIDT which is a nice palindrome by the way but which means top-down induction of decision trees we will look in more detail on a particular algorithm that been important for this area called the ID3 algorithm and we will discuss a few crucial issues around the application of that algorithm. The third lecture is about instance-based learning them meaning looking at learning techniques where we essentially store all instances but don't know build up any abstractions and essentially this lecture have two parts, the first part is around one kind of algorithm which is called the K nearest neighbour algorithm and we will discuss the properties of this kind of algorithm. And in the second part we will study a some machine learning schemes that are very important today because there been many successful applications, they are discussed here because with Logic tree did they go very well together yeah with the instance based learning approach, and actually the first part here is something called linear classifier which is the basic kind of machine learning strategy and then a particular form of linear classifier which is term support vector machines. Finally we will look into some specific techniques which call kernel methods that we will apply to the linear classifier in order for that kind of approach to also be able to handle nonlinear cases. The last lecture will be about Cluster Analysis and cluster analysis being a technique that need to be applied typically in the first stages of an unsupervised learning scenario, where we have to look at large sets of unclassified examples and essentially there are many approaches in this area I think it's been said there are more than hundreds of algorithms that try to solve the clustering problem and what we will do here is we will try to sort these all these approaches into five categories Partition-based, Hierarchical Based, Density based, Grid based and model based, and I will try to in a pretty summarized fashion give you a picture of the approach for each one of these categories. So this was the end of this introductory lecture, thanks for your attention we will continue with the next lecture on the topic of generalization as search thank you very much.