Global View of Meta-Learner Configuration Space

Goal

Discuss the surface features of the accuracy-metric surface in the unconstrained meta-learner N space. This N space is created by appending several different spacial realms together to incorporate all of the data for meta-learning into a single realm.

K Space (Option-Accuracy Space)

The first space is the individual meta-learner accuracy realm created from the K-1 options for all of the individual machine learners. The last variable is the accuracy of the meta-learner at a particular option setting. Measuring the meta-learner's accuracy on the Test Set for the entire range of option settings for this meta-learner configuration for a certain dataset will give a surface in K+1 Space. For ease we will call this surface S_K. We can look at a single run of S_K and talk about its Maximum point being the most accurate option setting for this meta-learner configuration and this data set. We could then compute this surface many times to get a “fuzzy” surface and discuss its variance (wrt individual dimensions) at various points in the option space. Now we can do the same thing for a completely new dataset. This gives us a completely different “fuzzy” surface (or does it) for this dataset. We can then look at the extrema points of this surface and compare it to the extrema of the surface for the first dataset. Perhaps we can look at features of the individual datasets (big or small, strong data correlation) to make some inference about the extrema. This brings us to the next realm space.

M Space (Metric Space)

The M space is created from the M metrics that we have for evaluating individual datasets. These metrics are assumed to be continuous. For every dataset we get a single point in this space. Now if we plot all of our datasets we can talk about their distribution in M-Space and other features of the point cloud. How can we describe this point cloud? How about sub-areas of this cloud?

L Space (Metric-Option Space)

Now we append the M space to the K space to get a new L space which we'll call the Metric-Option Space. Now when we look at the accuracy surface in L space (S_L) and we have a surface that contains not only the meta-learner's accuracy on the dataset, but also the metric data about the dataset on which the machine learners were trained. Now we can look at the surfaces from several different datasets and talk about their features. Do surfaces from similar datasets behave in a similar manner? Can we look at an unknown dataset and determine from its position in M-Space how the options should be set for a quasi-optimal meta-learner training (note that this is only for this particular Meta-Learner configuration). What do surface intersections mean in tis space? What does variance mean in this space?

C Space (Meta-Learner Configuration Space)

A individual L space contains only the results for a single meta-learner configuration. This means that for each possible Meta-Learner configuration (and there are infinite configurations) has a discrete point in space. Caruana dealt with creating a super meta-learner by subsamples of this space. Each “bin” in C-Space is a separate L-Space. Now we can look at a huge surface (let's call it a Super Surface) in the C-Space which is the collection of possible L-Space surfaces (note that this super surface will likely have many discontinuities). Now we can find the Super Surface for a particular dataset and if we can find the absolute maxima (not an easy task at all) then we have the single point (assuming that there is a single maxima) which corresponds to the most accurate meta-learner configuration possible. Now we could do this for many datasets and perhaps find a Super Meta-Learner, which will be ideal for most datasets. More likely we can find good starting points for similar datasets such that if given an unknown dataset we can known which meta-learner configuration is likely to produce a VERY good result.

Operational Complexity

The problem with this approach is that it is impossible to explore an infinite multidimensional space in its entirety, but we can explore various areas and from that try to extrapolate the features of the space we haven't explored. One approach would be to implement L-space gradient ascent. This would allow for a more efficient exploration of the various C-space bins since you wouldn't have to compute every point in the L-space. Rather, you could use gradient ascent to find the maxima in each space, which is the real information that we are interested in.

Class Exploration

If multiple researchers pick different regions to explore then perhaps we can get enough data to make decisions about better places to search for more global maxima.