Explicit Utility in Supervised Learning

James L. Carroll, Neil Toronto, Robbie Haertel, and Kevin D. Seppi

Primary Contact: James L. Carroll (jlcarroll@gmail.com) http://aml.cs.byu.edu

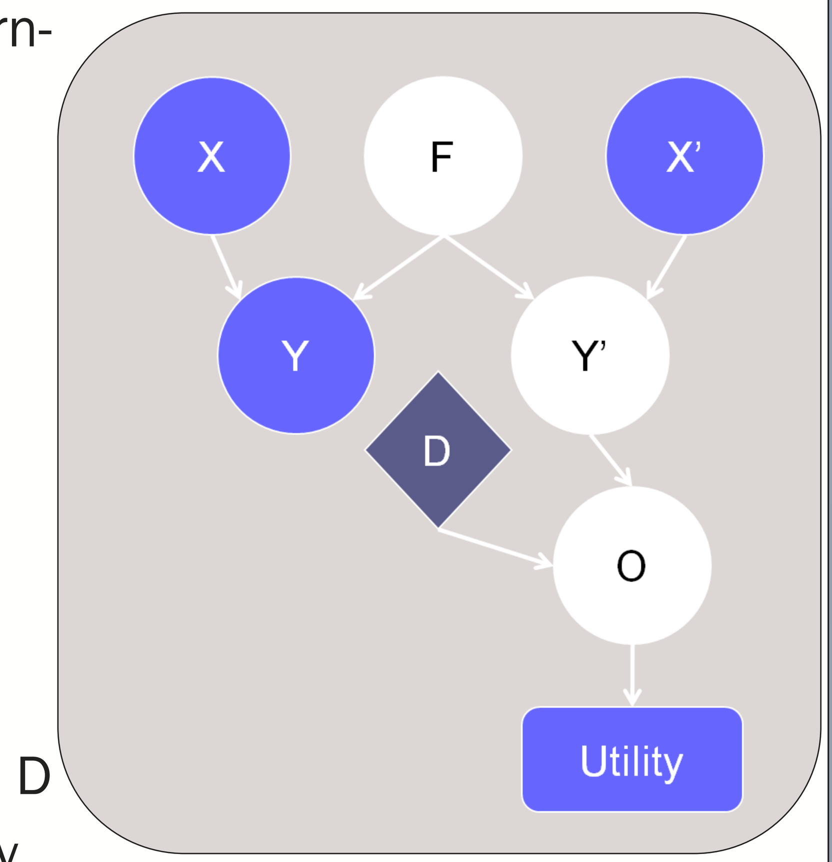
Introduction

We use a graphical model of the supervised learning problem together with the principles of decision theory to explore the theoretical effect of utility on supervised learning, No-Free-Lunch, sample complexity, and active learning, where utility is defined as sample costs and end-use utilities.

UBDTM Model of the Supervised Learning Problem

The Utility Based Decision Theoretical Model of supervised learning can be posed as follows:

- X: Feature Inputs,
- Y: Classification Outputs,
- F: The unknown function that maps inputs to outputs,
- X and Y are in the training set,
- X' and Y' are in the test set,
- O is the outcome of decision D which leads to an end-use Utility.



Sources of Utility/Costs

There are two sources of Utility in the Supervised Learning Problem:

- Utility from end-use,
- Negative utility from sample cost.

Optimal Supervised Learning

$$p(f\,|\,x,y)=\frac{p(y\,|\,x,f)p(f)}{\int p(y\,|\,x,f)p(f)df}$$

$$p(y'|\,x',x,y)=\int p(y'|\,x',f)p(f\,|\,x,y)df$$

$$\hat{d}=\underset{d\in D}{\arg\max}\sum_{y'\in Y}U(o)p(o\,|\,y',d)p(y'|\,x',x,y)$$

Separating Utility Free and Utility Dependant Learning

- The probability of a class given features and training data p(y'|x',TrainingData) can be computed independently of both sources of utility.
- The optimal decision given a training set can be computed independently of sample cost but not independently of end-use utility.
- Traditional theory tends to conflate Y' and D. By separating Y' from D, the effects of utility are more easily analyzed.
- The optimal decision given an arbitrary end-use utility function, can only be computed given the full posterior predictive distribution p(y'|x',TrainingData) .
- If all end-use utility functions are equally likely (if we really don't know how our algorithm will be used) then all decisions have equal expected utility, and all algorithms therefore perform equally well.

Utility Implications for No-Free-Lunch (NFL)

Within the context of the UBDTM model, NFL can be restated as:

- A uniform prior over F leads to all decisions D having the same expected utility off training set for the 0-1 loss end-use utility function.

A novel utility-free version of NFL can now be formulated:

- A uniform prior over F leads to a uniform posterior predictive over Y' off training set for all end-use utility functions,
- "A-priori" distinctions between supervised learning algorithms are possible for some end-use utility functions.
  - —For some utility functions, supervised learning algorithms that do not return the uniform posterior predictive demanded by the laws of Bayesian Statistics can lead to sub-optimal results.

Utility Implications for Sample Complexity

Sample complexity should depend on both utility from end-use and on utility from sample cost.

- Traditional sample complexity techniques assume uniform sample cost and 0-1 loss end-use utility.
- When end-use requires higher accuracy, sample complexity should be higher.
- When sample costs are higher, sample complexity as measured by costs should be higher.
  - — This effect is especially important when sample costs are not uniform.

Utility Implications for Active Learning

Active learning depends on both utility from end-use and on utility from sample cost.

- For example, if end-use does not depend on differentiating between some classes then then active learning should not focus on learning to distinguish between those classes.
- For example, if some samples are more expensive than others, then active learning should find the cheapest samples that still provide the most value of information.

One real world situation where sample costs can change is based on part of speech annotation where there are many possible scenarios for paying annotators. In this task, the performance of active learners actually swaps across random as the technique for paying annotators changes.

AL/Cost	Best	Middle	Worst
Pay by Sentence	Longest	Rand	QBU
Pay by Word	QBU	Rand	Longest

Conclusion: Advice for Practitioners

- Common applications for supervised learning often involve more complex end-use utility functions than 0-1 loss with more complex sample costs functions than uniform.
- End-use utility assumptions affect NFL, Sample Complexity, and Active Learning.
  - —Classifiers that report full posterior predictives are flexible to changes to end-use.
- Sample cost utility assumptions affect Sample Complexity and Active Learning.
  - —New techniques for Sample Complexity and Active Learning should be developed which are sensitive to end-use and sample costs.