Support Vector Machines for Differential Prediction

Finn Kuusisto[1], Vitor Santos Costa[2], Houssam Nassif[3], Elizabeth S. Burnside[1], David Page[1], and Jude Shavlik[1]

[1]University of Wisconsin, Madison, USA [2]University of Porto, Porto, Portugal [3]Amazon, Seattle, USA

Overview

Differential prediction has broad and important applications across a range of domains and, as specific motivating applications, we will consider two medical tasks. One is a task in which we want to specifically identify older patients with breast cancer who are good candidates for "watchful waiting" as opposed to treatment. The other is a task in which we want to specifically identify patients who are most susceptible to adverse effects of COX-2 inhibitors, and thus not prescribe such drugs for these patients.

Adverse COX-2 Inhibitor Effects Task

- Non-steroidal anti-inflammatory drug (NSAID)
- Significantly reduced occurrence of adverse gastrointestinal effects common to other NSAIDs (e.g. ibuprofen)
- Rapid and widespread acceptance for treatment of ailments such as arthritis
- Clinical trials showed significant increase in risk of myocardial infarction (MI), or "heart attack"

Identify patients susceptible to an increased risk of MI as a direct result of taking COX-2 inhibitors.

In Situ Breast Cancer Task

- Most common cancer in women
- Two basic stages: In situ and invasive
  - *In situ* cancer cells are localized
  - *Invasive* cancer cells have infiltrated surrounding tissue
- Younger women have aggressive in situ cancer
- Older women often have indolent in situ cancer

Identify older patients with in situ breast cancer that is distinct from that of younger patients.

Differential prediction requires the ability to measure differences in classifier performance between two subgroups. The standard measure of differential prediction in marketing is uplift, which is defined as the difference between the lift for the two subgroups. We propose and implement the SVM^{Upl} model, which optimizes uplift directly, and obtain excellent results.

Selected References

Hansotia, B., Rukstales, B.: Incremental value modeling. Journal of Interactive Marketing (2002).

Joachims, T.: A support vector method for multivariate performance measuers. In: Proceedings of the 22nd International Conference on Machine Learning (2005).

Nassif, H., Kuusisto, F., Burnside, E., Page, D., Shavlik, J., Santos Costa, V.: Score as you lift (SAYL): A statistical relational learning approach to uplift modeling. In: European Conference on Machine Learning (2013).

Radcliffe, N. and Simpson. R.: Identifying who can be saved and who will be driven away by retention activity. Journal of Telecommunications Management (2008).

Rzepakowski, P., Jaroszewicz, S.: Decision trees for uplift modeling with single and multiple treatments. Knowledge and Information Systems (2012).

Tuffery, S.: Data Mining and Statistics for Decision Making. John Wiley & Sons, 2nd edn. (2011).

Zaniewicz, L., Jaroszewicz, S.: Support vector machines for uplift modeling. In: IEEE ICDM Workshop on Causal Discover (2013).

Uplift Modeling

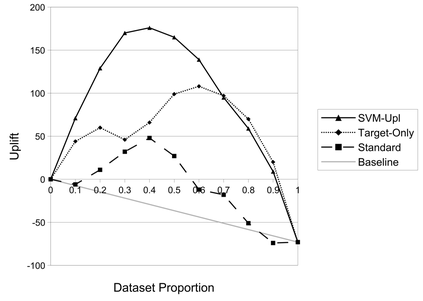
The reference work in uplift modeling originates from the marketing domain, in which customers can be broken into four *latent* categories:

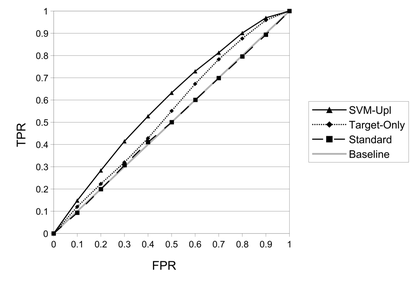
Persuadables Customers who respond positively (e.g. buy a product) when targeted.
Sure Things Customers who respond positively regardless of being targeted.
Lost Causes Customers who do not respond (e.g. not buy a product) regardless of being targeted.
Sleeping Dogs Customers who do not respond as a result of being targeted.

Ideally, only Persuadables would be targeted, but it's impossible to know any particular customer's persuadability as they cannot be both targeted and not targeted by some marketing activity. Only the eventual response and target/control status of an individual is known. Uplift modeling tries to tease out the latent information, relying on the uplift measure for evaluation.

Target		Control	
Response	No Response	Response	No Response
Persuadables, Sure Things	Sleeping Dogs, Lost Causes	Sleeping Dogs, Sure Things	Persuadables, Lost Causes

Does it work?
We generated a synthetic customer population and simulated marketing activity such that we knew the ground truth customer groups. We evaluated the trained models using the standard uplift measure, and with an ROC curve where the true Persuadables were treated as the positive class and all other categories were negative.





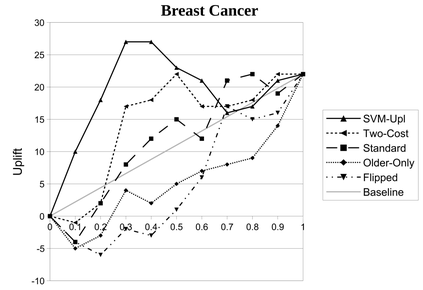
How does it apply to medical tasks?
In the COX-2 inhibitor task, variability in response to the drug suggests that there will be some people at increased risk of MI as a result of taking the drug, some who are at increased risk regardless, some who are at decreased risk regardless, and so on. Like in the marketing task, an individual cannot both take the drug and not take the drug to determine its effect. We propose that training a classifier to identify individuals for which taking a COX-2 inhibitor increases their risk of MI is analogous to identifying Persuadables. In the breast cancer task, the analogy is not as obvious, but we know that younger patients often have aggressive cancers while older patients have both aggressive and indolent cancers. Again, which type of cancer a patient has is not directly observable and it is unreasonable to not treat patients in an attempt to determine which have less aggressive varieties. We propose that training a classifier to identify less aggressive varieties of cancer (seen in older patients) is also analogous to identifying Persuadables.

Results

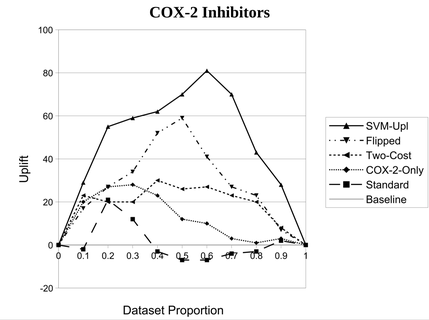
We used 10-fold cross-validation for evaluation. Final curves were produced by merging the output test results for each fold. Cost parameters were selected for each fold using 9-fold internal cross-validation. For all approaches, the cost parameter was selected from 1.0×10^{1} through 1.0×10^{-5}. For the two-cost model, C_A and C_B were selected from all combinations of the values such that $C_A > C_B$. We use the Mann-Whitney test at the 95% confidence level to compare approaches to SVM^{Upl} based on per-fold AUU (* indicates significance).

Older		Younger	
In Situ	Invasive	In Situ	Invasive
132	401	110	264

COX-2		No COX-2	
MI	No MI	MI	No MI
184	1,776	184	1,776



Model	Older AUL	Younger AUL	AUU	Per-fold AUU μ	Per-fold AUU σ	SVM^{Upl} p-value
SVM^{Upl}	64.26	45.05	19.21	1.93	0.78	-
Two-Cost	74.30	60.76	13.54	1.45	1.18	0.432
Older-Only	67.70	61.85	5.85	1.03	1.15	0.037*
Standard	75.35	64.34	11.01	1.26	0.38	0.049*
Flipped	53.90	49.08	4.82	0.77	0.58	0.020*
Baseline	66.00	55.00	11.00	1.10	0.21	0.004*



Model	COX-2 AUL	No COX-2 AUL	AUU	Per-fold AUU μ	Per-fold AUU σ	SVM^{Upl} p-value
SVM^{Upl}	123.38	72.70	50.68	5.07	2.04	-
Two-Cost	126.23	106.25	19.99	2.43	1.54	0.004*
COX-2-Only	151.50	137.70	13.80	1.18	1.52	0.002*
Standard	147.69	146.49	1.20	-0.16	1.25	0.002*
Flipped	102.15	73.63	28.52	2.97	1.35	0.037*
Baseline	0.00	0.00	0.00	0.00	0.00	0.002*

Uplift

The fundamental property of differential prediction is the ability to quantify the difference between the classification of subgroups in a population.

Lift

The number of **true positives** that a classifier achieves **at a given proportion of all examples labeled positive.**

Uplift

The **difference in lift** produced by a classifier between subgroups A and B **at a given proportion of all examples labeled positive.**

$$AUU = AUL_A - AUL_B$$

Maximizing Uplift

We rely on the relationship between AUC and AUL to extend work on support vector machines designed to maximize AUC into the uplift modeling domain. If we define the positive skew of data as $\pi = \frac{P}{P+N}$, then AUL is related to AUC by:

$$AUL = P\left(\frac{\pi}{2} + (1-\pi)AUC\right)$$

Using this, AUU is related to AUC by:

$$AUU = P_A\left(\frac{\pi_A}{2} + (1-\pi_A)AUC_A\right) - P_B\left(\frac{\pi_B}{2} + (1-\pi_B)AUC_B\right)$$

Furthermore:

$$max(AUU) \equiv max(P_A(1-\pi_A)AUC_A - P_B(1-\pi_B)AUC_B)$$
$$\propto max\left(AUC_A - \frac{P_B(1-\pi_B)}{P_A(1-\pi_A)}AUC_B\right)$$

Defining $\lambda = \frac{P_B(1-\pi_B)}{P_A(1-\pi_A)}$, we have:

$$max(AUU) \equiv max(AUC_A - \lambda AUC_B)$$

Further details can be found in the paper.

Conclusions

We introduced a support vector model directed toward differential prediction. The SVM^{Upl} approach optimizes uplift by relying on the relationship between AUL and AUC, and on the linearity of the multivariate function used in prior work to optimize AUC. The results suggest that SVM^{Upl} does indeed achieve better uplift in unseen data than the other approaches.

Acknowledgments

This work is supported by NIH grant R01-CA165229, NIGMS grant R01-GM097618, and NLM grant R01-LM010921. VSC was funded by the ERDF through the Progr. COMPETE, the Portuguese Gov. through FCT, proj. ABLe ref. PTDC/EEI-SII/2094/2012, ADE (PTDC/ EIA-EIA/121686/2010).